

Data's shameful neglect

Research cannot flourish if data are not preserved and made accessible. All concerned must act accordingly.

More and more often these days, a research project's success is measured not just by the publications it produces, but also by the data it makes available to the wider community. Pioneering archives such as GenBank have demonstrated just how powerful such legacy data sets can be for generating new discoveries — especially when data are combined from many laboratories and analysed in ways that the original researchers could not have anticipated.

All but a handful of disciplines still lack the technical, institutional and cultural frameworks required to support such open data access (see pages 168 and 171) — leading to a scandalous shortfall in the sharing of data by researchers (see page 160). This deficiency urgently needs to be addressed by funders, universities and the researchers themselves.

Research funding agencies need to recognize that preservation of and access to digital data are central to their mission, and need to be supported accordingly. Organizations in the United Kingdom, for instance, have made a good start. The Joint Information Systems Committee, established by the seven UK research councils in 1993, has made data-sharing a priority, and has helped to establish a Digital Curation Centre, headquartered at the University of Edinburgh, to be a national focus for research and development into data issues. Other European agencies have also pursued initiatives.

The United States, by contrast, is playing catch-up. Since 2005, a 29-member Interagency Working Group on Digital Data has been trying to get US funding agencies to develop plans for how they will support data archiving — and just as importantly, to develop policies on what data should and should not be preserved, and what exceptions should be made for reasons such as patient privacy. Some agencies have taken the lead in doing so; many more are hanging back. They should all be moving forwards vigorously.

What is more, funding agencies and researchers alike must ensure that they support not only the hardware needed to store the data, but

also the software that will help investigators to do this. One important facet is metadata management software: tools that streamline the tedious process of annotating data with a description of what the bits mean, which instrument collected them, which algorithms have been used to process them and so on — information that is essential if other scientists are to reuse the data effectively.

Also necessary, especially in an era when data can be mixed and combined in unanticipated ways, is software that can keep track of which pieces of data came from whom. Such systems are essential if tenure and promotion committees are ever to give credit — as they should — to candidates' track-record of data contribution.

Who should host these data? Agencies and the research community together need to create the digital equivalent of libraries: institutions that can take responsibility for preserving digital data and making them accessible over the long term. The university research libraries themselves are obvious candidates to assume this role. But whoever takes it on, data preservation will require robust, long-term funding. One potentially helpful initiative is the US National Science Foundation's DataNet programme, in which researchers are exploring financial mechanisms such as subscription services and membership fees.

Finally, universities and individual disciplines need to undertake a vigorous programme of education and outreach about data. Consider, for example, that most university science students get a reasonably good grounding in statistics. But their studies rarely include anything about information management — a discipline that encompasses the entire life cycle of data, from how they are acquired and stored to how they are organized, retrieved and maintained over time. That needs to change: data management should be woven into every course in science, as one of the foundations of knowledge. ■

"Data management should be woven into every course in science."

A step too far?

The Obama administration must fund human space flight adequately, or stop speaking of 'exploration'.

After the space shuttle *Columbia* burned up during re-entry into Earth's atmosphere in 2003, the board that was convened to investigate the disaster looked beyond its technical causes to NASA's organizational malaise. For decades, the board pointed out, the shuttle programme had been trying to do too much with too little money. NASA desperately needed a clearer vision and a better-defined mission for human space flight.

The next year, then-President George W. Bush attempted to supply that vision with a new long-term goal: first send astronauts to build

a base on the Moon, then send them to Mars. This idea immediately set off a debate that is still continuing, in which sceptics ask whether there is any point in returning to the Moon nearly half a century after the first landings. Why not go to Mars directly, or visit near-Earth asteroids, or send people to service telescopes in the deep space beyond Earth?

Yet that debate is both counter-productive — a new set of rockets could go to all of these places — and moot, because Bush's vision never attracted the hoped-for budget increases. Indeed, a blue-ribbon commission reporting to US President Barack Obama this week (see page 153) finds the organizational malaise unchanged: NASA is still doing too much with too little. Without more money, the agency won't be sending people anywhere beyond the International Space Station, which resides in low Earth orbit only 350 kilometres up. And even the ability to do that is in question: Ares I, the US rocket that would return

astronauts to the station, is potentially on the chopping block.

NASA critics can rightly point out that the benefits of human space flight are fuzzy, especially when it comes to the science. The returns are occasionally bountiful, as with the astronauts' recent repair of the Hubble Space Telescope. But for the most part they are incidental and hugely expensive.

NASA-funded space scientists might be excused for feeling a bit smug. Their robotic science missions to Mars and elsewhere are orders of magnitude more cost-effective. And their budgets remain relatively protected from the turmoil engulfing the debate on human space flight — as they should be. Indeed, Obama's budget proposals bolster NASA's Earth-observation programme, where some of the most pressing knowledge is to be gained.

Like it or not, however, scientists do have a stake in the human space-flight debate. The rockets and the technology developed to take astronauts beyond Earth orbit could also make it possible to

mount much more ambitious robotic missions. And perhaps even more important, the sight of humans travelling beyond Earth has an undeniable power to inspire future generations of space scientists (see *Nature* **460**, 314–315; 2009). This link should not be surprising: both endeavours are animated by the same spirit of exploration.

True, sending astronauts beyond low Earth orbit is never going to be cheap. But adequately funding the 2004 exploration vision would not require money on the scale of the Manhattan Project, or even the Apollo programme. A boost of a few billion dollars a year — perhaps 15% of NASA's \$17.6-billion total budget — would allow the agency to pursue a long-term programme of heavy-lift rockets that could go to the Moon, or other deep-space locales.

If Obama is not willing to support such a plan, then he and the American public should stop pretending that they are in favour of human space exploration. Because maintaining the space station is not exploration. It is a commute. ■

Overrated ratings

Criteria for 'green buildings' need to make energy performance a priority — as do universities.

The American College and University Presidents' Climate Commitment, a pledge by some 650 US institutions of higher education to eventually make their campuses carbon neutral (see page 154), is an effort that should be encouraged and expanded. Buildings account for an estimated 45% of the world's total energy consumption and a similar share of its greenhouse-gas emissions; the classrooms, laboratories and other structures in US universities collectively generate some 42 million tonnes of carbon dioxide per year.

However, one emissions-reduction mechanism endorsed by the commitment deserves a more sceptical look than it often gets. This is a requirement that all new campus structures aim for certification under the Leadership in Energy and Environmental Design (LEED) rating scheme developed by the US Green Building Council (USGBC).

LEED is the best known of several internationally recognized rankings for environmentally conscious design. Launched in 1998, it now encompasses 14,000 projects in the United States and 30 other countries. Yet, as is well known in the building research community but not outside it, neither LEED nor any other such rating is a reliable guide to energy performance. Labelled buildings often perform no better in energy terms than the general building stock, and sometimes worse.

One reason is that the energy performance is not the only measure used in the ratings. LEED, for example, also awards greenness points for the choice of a site that protects the environment and wildlife; the use of sustainable, environmentally friendly materials; water and waste management; and indoor air quality. Another reason is that most ratings assess a building's energy performance using theoretical projections from engineers' models, but don't measure its real, post-occupancy performance, which often can be much poorer.

Issues of indoor environmental quality and sustainability are important. But given the urgency of addressing climate change —

plus the fact that a high green-building rating is often taken to be an energy certification, even when it is not — the schemes should give energy performance considerably more priority than they have to date (see *Nature* **452**, 520–523; 2008).

In April, the USGBC took a welcome step in that direction, releasing a revised version of its scoring system that gives energy performance more weight. And this month it announced an equally welcome initiative to collect post-occupancy data, while carrying out research with academic partners to better compare these data with predicted performances. This is an area that, like most green-building research, has been abysmally underfunded in the past.

If universities wish to set an example in climate-change efforts, they too must place greater emphasis on building-energy performance.

One way to accomplish this would be to supplement green-building ratings, such as LEED, with dedicated energy-performance ratings, such as the Swiss Minergie standard, which focuses exclusively on the bottom line: a building's annual energy consumption per square metre.

By setting higher standards than local government regulations, voluntary rating systems such as LEED have undeniably raised public awareness of sustainable building practices, and have stimulated the adoption of those practices across the building profession. Despite this, progress in reducing the energy consumption of buildings remains negligible compared with its huge potential for reducing global CO₂ emissions.

Likewise, the US colleges and universities that have signed up to the climate commitment have done the right thing by setting their own energy performance bar high enough to inspire other organizations, and to help stimulate broader change across the economy. But, as former US president Bill Clinton said last month at a summit meeting of the commitment in Chicago, Illinois: "For all the good we're doing, we're just piddling compared to what we ought to be doing, and compared to what we could be doing." ■

"For all the good we're doing, we're just piddling compared to what we ought to be doing, and compared to what we could be doing."

RESEARCH HIGHLIGHTS



G. DABB

Warning wings

Proc. R. Soc. B doi:10.1098/rspb.2009.1110 (2009)

The sound made by feathers may make for a useful warning signal when birds flock together.

Mae Hingee and Robert Magrath at the Australian National University in Canberra studied the crested pigeon, *Ocyphaps lophotes* (pictured), which makes a fluttering metallic

sound when it flaps its wings. From recordings, they found differences between the sounds of this wing 'whistle' during normal take-offs and those of panicked flights made in response to a threat.

They then played back the sounds to groups of pigeons. Calm take-offs had no effect, but recordings of alarmed birds frequently sent flocks scattering.

ATMOSPHERIC CHEMISTRY**Ozone's winners and losers**

Nature Geosci. doi:10.1038/ngeo604 (2009)

Changes in atmospheric circulation driven by global warming could shift the global distribution of ozone northwards.

Michaela Hegglin and Theodore Shepherd at the University of Toronto in Canada isolated the effects of global warming by simulating ozone interactions in an atmospheric chemistry climate model. They focused on the decades 1960–70 and 2090–2100, representing periods before and after the most severe effects of ozone-depleting chemicals. Climate change increases tropical upwelling, pushing ozone into northern latitudes. At the same time, southern latitudes see a decrease in ozone transport.

As a result, the authors report that by the end of this century ultraviolet radiation could decrease by 9% at high northern latitudes; tropical regions could see an increase of 4%; and southern high latitudes could receive up to 20% more in the late spring and early summer.

PHYSICS**Magnetic monopoles**

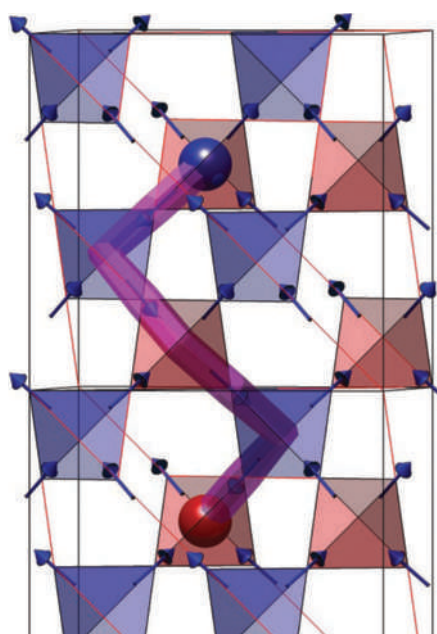
Science doi:10.1126/science.1178868; 10.1126/science.1177582 (2009)

Physicists have searched for decades for magnets with a single pole. Now, two independent groups report the latest signatures of magnetic monopoles in a class of crystalline materials called spin ice. When the crystals were chilled to near absolute zero, they seemed to fill with tiny single points of north and south separated by fractions

of a nanometre. Jonathan Morris at the Helmholtz Centre for Materials and Energy in Berlin and his colleagues used $\text{Dy}_2\text{Ti}_2\text{O}_7$, whereas Tom Fennell at the Institute Laue–Langevin in Grenoble, France, and his collaborators used $\text{Ho}_2\text{Ti}_2\text{O}_7$.

The atoms in the crystals sit at the corners of tetrahedra. Each atom behaves like a tiny bar magnet, and when the crystal is cooled, the atoms align to create regions of north or south magnetic charge, separated by a chain of aligned atoms (see image, below). The charge isn't attached to any physical object, but it behaves like a monopole.

For a longer story on this research, see <http://tinyurl.com/monopole>

**COMPUTATIONAL BIOLOGY****A new protein subdivision**

Cell 138, 774–786 (2009)

The traditional hierarchy of protein structure might require revision. Rama Ranganathan at the University of Texas Southwestern Medical Center in Dallas and his colleagues propose that proteins contain semi-independent clusters of co-evolving amino acids that they call 'protein sectors'.

The researchers analysed the conserved biological properties of the S1A protein family. They found that the proteins' amino acids are organized into three conserved functional units that are distinct from the classically observed structural hierarchies based on sequence or three-dimensional shape. As such, the authors argue, natural selection may operate at the level of these protein sectors.

MICROBIAL EVOLUTION**Cholera gene swap**

Proc. Natl Acad. Sci. USA 106, 15442–15447 (2009)

Cholera has affected humans for more than a hundred years, but how the bacterium that causes the disease, *Vibrio cholerae*, has evolved had not been described.

Rita Colwell of the University of Maryland in College Park and her collaborators compared the genomes of 23 strains of the bacteria isolated over the past 98 years. They found that the strains responsible for the current cholera pandemic, which started in 1961, are descendants of a single strain, and evolved mainly through gene transfer with other strains in the environment. The culprits behind the previous pandemic, in the early

SCIENCE/AAAS

twentieth century, were from a different lineage altogether.

Because of the bacteria's rapidly evolving genome, the researchers say that cholera strains should be identified by gene content rather than by cell-surface protein marker.

CHEMISTRY

Going for gold

Nano Lett. doi:10.1021/nl902186v (2009)

To increase conductance in miniaturized circuits, just add gold.

Paul Alivisatos at the University of California, Berkeley, and his colleagues were looking to see how the interface between a semiconductor nanorod and a metal would affect conductance. So they immersed the 40-nanometre-long cadmium selenide rods in a solution containing gold. This capped the rod tips with gold directly, and avoided the formation of a gold–semiconductor alloy, or a surfactant layer on the nanorod tip — both consequences of other rod-making procedures.

The procedure placed gold atoms on the tips of the rods and decreased the barrier to conductance — known as a Schottky barrier — giving them 100,000 times improved behaviour.

EVOLUTION AND DEVELOPMENT

Genes in the mirror

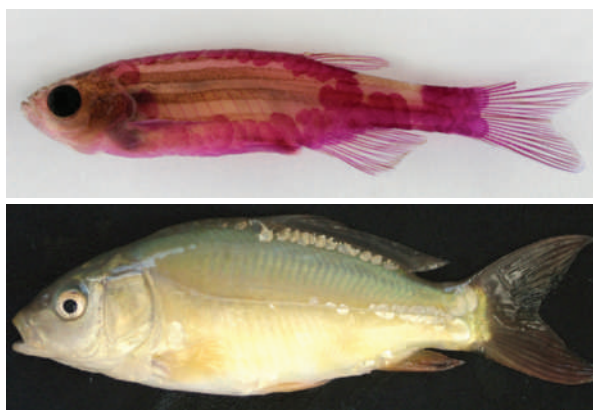
Curr. Biol. doi:10.1016/j.cub.2009.07.065 (2009)

Gene duplications may serve as the raw material for evolutionary changes in physical traits. Now researchers have found an example of how this might work in mutant and domesticated fish.

Nicolas Rohner and Matthew Harris of the Max Planck Institute for Developmental

Biology in Tübingen, Germany, and their co-workers created a zebrafish (*Danio rerio*, pictured in upper panel) that is mostly scaleless. They found that the mutation responsible is in the gene *fgfr1*, which is required for embryonic development. Scouring the fish's genome, the authors found a previously uncharacterized copy of the gene. The two compensate for one another during embryonic growth but control different traits in adulthood.

Mutations in the same gene cause the characteristic scale pattern seen in the domesticated mirror carp (*Cyprinus carpio*, pictured in lower panel), providing a real-world example of how gene duplication can form the basis of new traits.



NEUROSCIENCE

Fear net

Science 325, 1258–1261 (2009)

When adult rats are trained to associate a sound with an electric shock, they will often fear that sound for a lifetime. Young rats, however, can erase the fear memory when it is no longer relevant.

Andreas Lüthi, Cyril Herry and their

colleagues at the Friedrich Miescher Institute for Biomedical Research in Basel, Switzerland, found that the reason for the shift may be the development of the perineuronal net, an extracellular protein lattice that surrounds a subset of neurons as rats mature. When the authors used an enzyme to dissolve the perineuronal net in the amygdala, a crucial brain region for forming fear memories, they found that adult rats could wipe away the memory of the shock as if they were young.

GENETICS

Why Y knots

Cell 138, 855–869 (2009)

Theory predicts the chipping away of the male Y chromosome owing to the fact that, for the most part, it has no recombination partner during meiosis, the sexual form of cell division. But male-specific genes on the Y persist, protected in part by palindromic DNA repeats that are maintained through recombination events between each other. These repeats, say David Page of the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, and his collaborators, can also be Y's Achilles' heel.

The researchers propose a mechanism by which a crossover event after chromosomal replication links the two copies of the chromosome together at a palindrome, creating a larger, abnormal chromosome. These 'isodicentric' Y chromosomes are implicated in sex reversal, Turner's syndrome and male infertility. In a study of samples from 2,380 patients with suspected Y-chromosome defects, the authors identified 51 that apparently formed by this mechanism.

N. ROHNER

M. BERCSÉNYI

JOURNAL CLUB

Elena B. Pasquale
Burnham Institute for Medical Research, La Jolla, California

A biologist is gratified to find reconciliation for a conflicted receptor.

When giving talks on the involvement of the Eph family of receptor tyrosine kinases in cancer, I sometimes include a slide of the two-faced Roman god, Janus, to signify the dichotomies of Eph function in cancer cells. Most proteins have a clear-cut

function. Some 'moonlighting' proteins carry out two unrelated functions. It is, however, rare for a protein to toggle between opposing activities. The Eph receptors are proving to be such outliers.

High expression of Eph receptors has been correlated with a poor cancer prognosis, but so has Eph silencing. Accordingly, there is good evidence that the Eph receptors can promote as well as inhibit tumour development. In a reconciliation reminiscent of Hegelian synthesis, a recent paper begins to explain how the EphA2

receptor can both promote and inhibit cancer cells' migratory and invasive abilities.

EphA2 activation by ephrin ligands seems to be minimal in most types of cancer cell. Hui Miao and Bingcheng Wang of Case Western Reserve University in Cleveland, Ohio, and their co-workers have shown that the protein Akt — which can be powerfully cancer-promoting — hijacks EphA2 by phosphorylating one of its serine residues, enabling its pro-metastatic activities (H. Miao *et al.* *Cancer Cell* 16, 9–20; 2009).

Remarkably, binding by the ephrin-A1 ligand erases this phosphorylation and transforms EphA2 into an anti-invasive molecule.

These findings lead to the counterintuitive proposition that we should encourage rather than inhibit EphA2's ligand-dependent function. It will be interesting to see whether analogous switches convert other Eph receptors between malignant and benign phenotypes.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

NEWS BRIEFING

● POLICY

Space exploration: NASA's human space-flight programme does not have enough money to fulfil its vision of building a Moon base or sending astronauts to Mars. The conclusion was due to be delivered to US President Barack Obama this week in a report from an expert panel led by former aerospace executive Norman Augustine. The report outlines a range of alternatives for NASA, including sustaining the International Space Station beyond its scheduled de-orbit in 2016, and cancelling the Ares I rocket that is being developed to carry astronauts to the Moon. For more, see page 153.

Climate data: Delegates representing 155 nations at the World Climate Conference in Geneva agreed on 3 September to set up a **global climate service** providing long-term forecasts to users ranging from national governments to individual farmers. Over the next four months, a task force set up by the World Meteorological Organization will work out the practicalities of the service. But some countries are baulking at the suggestion that they will need to supply the service with data, citing issues such as national security or commercial interests that would prevent disclosure. For more, see page 159.

Climate costs: The United Nations' World Economic and Social Survey 2009 estimated last week that the **developing world** would need between US\$500 billion and \$600 billion annually from rich nations — around 1% of their GDP — to shift to cleaner energy and adapt to global warming. Even this amount, well above previous estimates, is dwarfed by a Chinese economic analysis. Environmental economists at Renmin University in Beijing suggest that if emissions in **China** are to peak by 2030,



K. DIANSEZIAN/GETTY IMAGES

TELESCOPE SAVED FROM FIRE

Californian firefighters saved the historic Mount Wilson observatory (above) from a massive arson fire that created huge palls of smoke and blackened more than 600 square kilometres in the mountains above Los Angeles. Astronomer Edwin Hubble used the 100-inch (2.5 metre) Hooker telescope on Mount Wilson in the 1920s to confirm that the Milky Way is one galaxy among many and that the Universe is expanding.

up to \$438 billion will have to be spent each year in that country alone.

Research ethics: Biomedical research collaborations between **Europe and China** need greater ethical oversight, according to BIONET, a panel that examines projects between the regions. At a meeting in London on 2–4 September, it recommended that a joint advisory body be set up to offer advice and monitor research practices in order to stamp out unregulated stem-cell therapies and prevent participants in clinical trials being exploited. For more, see page 157.

Emission projections: India's greenhouse-gas emissions will triple by 2031 but nevertheless will probably still be below the world per-capita average for 2005, said Jairam Ramesh, the country's environment minister, on 2 September. Citing five independent studies, he said emissions would rise from today's 1.2 billion tonnes to between 4 billion and 7.3 billion

tonnes of carbon dioxide equivalents — or between 2.77 and 5 tonnes per capita.

● BUSINESS

Pandemic flu: With flu season looming in the Northern Hemisphere, a small clinical trial by Novartis indicated that just one dose of pandemic H1N1 **flu vaccine** was sufficient to provoke an adequate immune response — if accompanied by an adjuvant, or booster chemical. If confirmed, this single-dose requirement would effectively double the amount of vaccine available, as two doses have been assumed necessary. Chinese company Sinovac said last month that a single shot of its vaccine also worked; it received a production licence from China's government last week.

Genome sequencing: The sequencing firm **Complete Genomics** in Mountain View, California, will not meet its goal this year of sequencing 1,000 human genomes for US\$5,000

SOUND BITES

“We’re walking through wet sand.”

Michael Zammit Cutajar, United Nations Framework Convention on Climate Change.

The chairman of a working group deliberating the text of a draft climate treaty for debate at December's Copenhagen summit describes the group's slow progress. (Reuters)

each. On 9 September, the company said it had sequenced just 14 genomes, for customers such as US drug giant Pfizer, and the HudsonAlpha Institute for Biotechnology in Huntsville, Alabama. Complete Genomics announced on 24 August that it was delaying its commercial launch by six months to January 2010, owing to fundraising difficulties.

Patent rejection: India's patent office has rejected claims from US companies Gilead and Tibotec for patents on their respective **HIV drugs**, tenofovir and darunavir. The decision opens the way for India to supply cheaper generic versions of the medicines, both to its own population and to other countries where the drug is not patented. It is the latest in a string of legal victories for Cipla, India's largest generic drug maker, which refused to sign up to a condition-bound licence on tenofovir that Gilead offered to generics manufacturers in 2006.

For a longer version of this story, see <http://tinyurl.com/lyqh8z>

Mineral shortage: Concern rose last week that China might further restrict exports of **rare earth minerals**, based on leaked details of a draft plan from the nation's Ministry of Industry and Information Technology. Wang Caifeng, deputy director-general of the ministry's Department of Raw Material Industry, told a mining conference in Beijing on 3 September that the policy was still under review, but insisted there would be no outright ban on exports of elements such as

NUMBER CRUNCH \$2.3bn

Record-setting sum paid by Pfizer to settle allegations that it had illegally marketed drugs and paid kickbacks to physicians.

dysprosium and terbium. China produces more than 90% of the world's rare earth elements, which are used as catalysts and in high-tech magnets, hybrid car batteries, wind turbines and mobile phones.

Mass spectrometry:

Leading mass-spectrometer manufacturer **AB SCIEX** has been bought for US\$1.1 billion by scientific and medical technology company **Danaher** of Washington DC. AB SCIEX was jointly owned by life-sciences companies Life Technologies of Carlsbad, California, and MDS of Mississauga, Ontario.

AWARDS

Balzan Prize: This year's winners, announced on 7 September, include two top scientists. **Michael Grätzel** of

the Federal Polytechnic School of Lausanne, France, was rewarded for his development of the dye-sensitized solar cell. **Brenda Milner** (pictured below), of McGill University in Montreal, Canada, received the prize for her research on the role of the hippocampus in the formation of memories. Awarded by the International Balzan Prize Foundation, the prize aims to bolster "initiatives in the cause of humanity, peace and brotherhood". The foundation, which is based in Milan, Italy, will give each of the winners 1 million Swiss francs (US\$944,000), half of which must be devoted to future research.



RESEARCH

Radio astronomy: The International Centre for Radio Astronomy Research opened last week in Perth, Australia, at a cost of Aus\$100 million (US\$85 million). The centre, largely funded by Curtin University of Technology and the University of Western Australia, both in Perth, is expected to help Australia's bid to host the Aus\$2.5-billion **Square Kilometre Array** radio telescope. A decision on whether Australia or its rival South Africa will host the array is expected in 2012.

THE WEEK AHEAD

13-18 SEPTEMBER

The European Planetary Science Congress holds its fourth annual meeting in Potsdam, Germany.

meetings.copernicus.org/epsc2009

14-18 SEPTEMBER

The International Atomic Energy Agency holds its annual general conference in Vienna.

www.iaea.org/About/Policy/GC

14-17 SEPTEMBER

Individual genomes' role in research and clinical medicine is the theme of Personal Genomes, the second meeting on the topic hosted by Cold Spring Harbor Laboratory, New York.

meetings.cshl.edu/meetings/person09.shtml

Geothermal halt: A flagship US\$17-million **geothermal energy** project has been halted after encountering problems at its northern California drilling site. AltaRock Energy of Sausalito, California, is backed by a \$6.25-million grant from the US Department of Energy and venture funding from Google. It aims to harness geothermal energy by cracking bedrock at the bottom of a deep well and pumping water through the cracks to generate steam. The company did not immediately provide details about the drilling problems, but is due to file a report with federal agencies.

BUSINESS WATCH

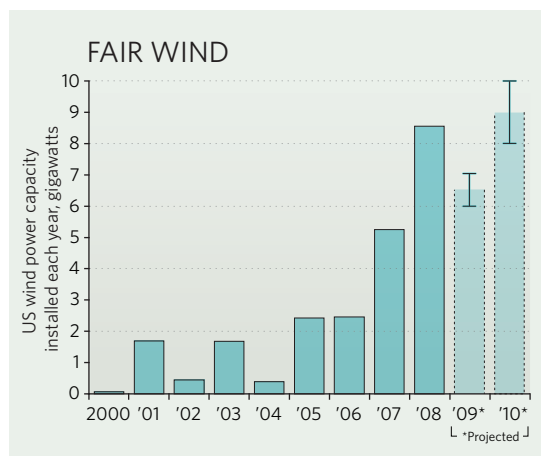
The US Department of Energy has released nearly \$500 million in direct cash subsidies for wind-energy developers, marking the first payment under a new stimulus programme intended to revive renewable-energy markets. The money will go to companies developing ten wind projects in six states; another \$3 million went to a pair of solar projects.

The cash payments are in lieu of tax incentives that have been in place for more than a decade. Energy developers previously secured up-front financing from banks, which then took advantage of the tax credits over time. That financing

dried up following the global economic downturn, contributing to a sharp decline in new wind-energy projects.

The US wind industry, previously hoping to install as much as 10 gigawatts of capacity in 2009, is now expecting around 6.5 gigawatts. But the London-based consultancy New Energy Finance is forecasting growth next year, with installation of between 8 and 10 gigawatts of capacity.

The energy department expects the new programme to provide around \$3 billion in funding, enabling projects valued at between \$10 billion and \$14 billion.



SOURCE: AWEA; PREDICTIONS FROM NEF

Cash crisis could ground NASA rocket

Crewed missions to the Moon are under threat, warns an expert panel.

A committee of aerospace engineers and scientists was poised to deliver its grim assessment of NASA's human space-flight programme to US President Barack Obama on 8 September. The panel's report will outline the stark choices Obama will face, which could include cancelling a new system of Moon-bound rockets and all but giving up on exploring space beyond the low Earth orbit of the International Space Station (ISS).

"The bottom line is, they concluded that there's not enough money in the current budget to do anything useful in human space flight," says Marcia Smith, president of the Space and Technology Policy Group, a consultancy based in Arlington, Virginia, and former director of the Space Studies Board at the US National Research Council.

In May, Obama ordered the committee to review the current space policy set by former president George W. Bush, with its "vision" of building a Moon base as a prelude to sending people to Mars. The committee was tasked with assessing new scenarios — including using the ISS past its scheduled de-orbit in 2016 — while keeping to strict budget guidelines. Led by former aerospace executive Norman Augustine, the ten-member committee has not yet released its report, but public discussions this summer have made some of the options clear.

Given the budget constraints, the choices weren't pretty. In Obama's 2010 budget request, NASA's exploration programme, known as Constellation, would receive about US\$6 billion per year — about \$1 billion less than Bush asked for in his 2009 budget, and several billion less than what was slated in previous budgets (see chart). "The Bush budget stressed the system, but the Obama budget, if left as is, breaks it," says Scott Pace, director of the Space Policy Institute at George Washington University in Washington DC. One analysis by the committee showed that if the current plan and budget are kept, astronauts won't even leave low Earth orbit until 2028.

So the panel looked at alternatives, narrowing down some 3,000 permutations to just a handful for presidential digestion. In several scenarios, the Ares I rocket — one of two needed to take cargo and astronauts to the Moon — would be cancelled. Instead, money would be poured into commercial space companies, such as Space Exploration Technologies of Hawthorne, California, and Orbital Sciences in Dulles, Virginia,



Current projects such as NASA's Ares I rocket could be cancelled in favour of commercial space flights.

which are already trying to build rockets to take cargo to the ISS. But the committee also seems inclined to support commercial rockets that could ferry people into space, says Smith.

Former NASA administrator Michael Griffin says there are risks not just in making a crewed commercial rocket a reality, but also in ceding the capability for space travel — traditionally held by the US government — to the private sector. "I am not a fan of attempts to rely on such a capability before it actually exists," says Griffin, now a professor of aerospace engineering at the University of Alabama in Huntsville. He says he would also be disappointed if Ares I were cancelled, not so much for the \$6 billion that has already been spent on the rocket and its *Orion* crew capsule, but because

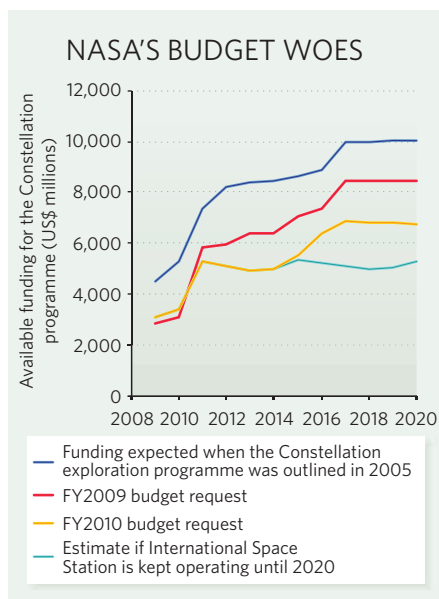
he still believes that Ares I is the cheapest way to get past low Earth orbit when paired with its heavy-lift launch companion Ares V. The system, he says, "has the sole failure of costing more than President Obama was willing to provide in the budget".

The committee found that extensive human exploration of the Moon and a direct trip to Mars are not feasible. With a little budgetary leeway, and with the Ares I money put into developing an alternative heavy-lift rocket, the committee determined that there could eventually be a 'deep space' option. Such possibilities could include visits to asteroids, flybys of the Moon and planets, and trips to Lagrangian points — the gravity wells in the Earth-Sun system where some telescopes are situated.

The committee found many ways to extend the operations of the ISS to 2020 in order to satisfy international agreements. What is not obvious is whether, after spending \$2.5 billion a year to service the ISS in coming years, there would be money for much else. "I dislike pretending that we have goals that are far-reaching and frontier-oriented when we're not willing to set aside money to achieve them," says Griffin.

Obama's 2010 budget guidance did include the caveat that additional money could be requested for the programme pending the Augustine committee's report. Congress, which is working to set those spending figures this autumn, has scheduled hearings on the report for mid-September. So although the committee's job will soon be over, some tough decisions — whether to argue for more money, or to accept a more limited programme — are still in store. "The more difficult job is going to be on the president's desk," says Smith.

Eric Hand



SOURCE: REVIEW OF HUMAN SPACE FLIGHT PLANS COMMITTEE

SPECIAL REPORT

How green is your campus?

Universities are working to bring sustainability to their campuses and classrooms, and could serve as a model for other institutions looking to go carbon-neutral. But there's no single way to grade the initiatives.

On a typically muggy day in late August, some 1,300 incoming freshmen and their parents gathered for orientation weekend at Emory University, near downtown Atlanta, Georgia. Here, in the heart of the conservative Deep South, the students received their first lesson of the school year. They were served food that was locally or sustainably produced, which they ate with cutlery made from sugar cane. And they were handed reusable water bottles and compact fluorescent light bulbs, which they toted around in reusable grocery bags. Over the two days of orientation, the school composted nearly two tonnes of waste, making it Emory's first near-zero-waste freshman orientation. "From the first time the students interact with Emory, we try to make it clear that sustainability is part of our DNA, that this is our expectation from them," says Ciannat Howett, director of the university's office of sustainability initiatives.

Emory is part of a wave of colleges and universities throughout the United States and across the globe that are going 'green'. "We've gotten into this situation where we have an unsustainable environmental future because we've produced all kinds of really smart people that don't get it," says Michael Crow, president of Arizona State University in Tempe. Crow is also chair of the American College & University Presidents' Climate Commitment, through which some 650 US educational institutions have pledged to become "climate neutral". Nearly 400 of them are now facing a 15 September deadline to submit their detailed 'climate action plans' for achieving their goals.

Measuring up

Such schools also hope to serve as models for others, including businesses, cities and counties, that hope to reduce their environmental impacts. But their experiences underscore the fact that sustainability can be hard to measure and that attaining it, especially with competing financial pressures, doesn't happen overnight.

More than 300 of the first signatories to the climate commitment have submitted greenhouse-gas inventories, which tally electricity use, heating and cooling of buildings, transportation to and from campus, and official air travel. Climate action plans are step two. So far, about 80% of the signatories have reported on time and are in good standing with the

initiative, says Anthony Cortese, president of the Boston-based non-profit organization Second Nature, which helps run the initiative. He expects 90% fulfilment by the beginning of the 2010–11 school year. Still, institutions set their own timetables for achieving climate neutrality, and there is no penalty if they fall short, aside from peer pressure by other members.

To quantify their greenhouse-gas reductions and efficiency gains, most schools rely on standardized emissions inventories, such as the Campus Carbon Calculator provided by Clean Air–Cool Planet, a non-profit group based in Portsmouth, New Hampshire. In some cases, institutions have their own environmental engineers or energy analysts who keep track of carbon accounting, with others engaging

are homing in on finding more centralized, lower-carbon alternatives for heating their buildings year-round.

Emissions gains

Some of the early starters have already made major advances in shrinking their carbon footprints and improving efficiency. Green Mountain College in Poultney, Vermont, is building a combined heat-and-power plant that will supply 85% of heating to the campus and run on renewable biomass such as locally sourced wood chips. Green Mountain's student enrolment has risen by 14% since 2007, but its carbon emissions per student have decreased by nearly 20%.

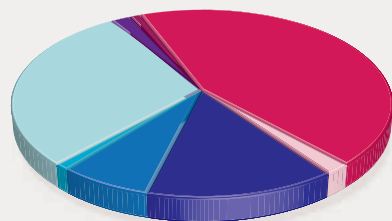
Meanwhile, the University of Minnesota, Morris, has constructed a large-scale wind-research turbine that supplies power to most of its buildings. And in 2008, Middlebury College in Vermont completed a biomass gasification plant, which is expected to replace 3.8 million litres of heating oil. Harvard University has more than 60 green building projects in progress. One of its building renovations, completed in 2008, resulted in a 35% improvement in energy efficiency and a 40% reduction in water use, says Heather Henriksen, the university's sustainability director.

And if the 51 institutions in one study succeed in going carbon-neutral, that would be equivalent to taking 690,000 cars off the roads, says Jason Pearlman of the consulting firm Sightlines, based in Guilford, Connecticut.

Some early sceptics, who once worried about universities trying to 'greenwash' their reputations with minor institutional adjustments, are now convinced. Dave Newport, director of the Environmental Center at the University of Colorado at Boulder, says that several years ago he was dubious about whether universities would really take a leading role in sustainability. "Campus leadership has really stepped up" since then, he says, "and the effort is nothing short of full speed ahead."

Many US schools have committed to meeting Leadership in Energy and Environmental Design (LEED) standards, set out by the US Green Building Council. In 2001, Emory built the first LEED-certified building in the south-east, a biomedical research building, and in 2005 it became the first US university to attain LEED certification for an existing building when it renovated its business school, a \$95,000

AVERAGE GROSS EMISSIONS, BY SOURCE, FOR US DOCTORATE-GRANTING UNIVERSITIES*



Stationary combustion	Purchased steam
Mobile combustion	Communting
Fugitive emissions	Air travel
Purchased electricity	Solid waste

* As reported by 85 signatories to the American College & University Presidents' Climate Commitment

students through their coursework. In addition, the Association for the Advancement of Sustainability in Higher Education, based in Lexington, Kentucky, has developed a system to help schools track their progress over time. Since February 2008, some 70 schools have piloted that system; it will officially launch in January, and its online reporting tool will be available to all campuses.

But it is difficult to find a universal system of ranking or grading sustainability, because schools grapple with different challenges, says David Oxtoby, president of Pomona College in Claremont, California. Whereas schools in the American West focus heavily on water conservation, for instance, many in New England

SOURCE: ACUPCC



Northern Arizona University is one of more than 600 colleges and universities that have signed up to an agreement to go 'green'.

NORTHERN ARIZONA UNIV.

project that paid for itself in less than a year through reduced energy bills, says Howett.

Institutions elsewhere are also jumping onboard. A junior college in Puerto Rico and a community college in the Republic of Palau have signed the climate commitment. Six educational institutions have also recently joined the Climate Neutral Network, led by the United Nations Environment Programme, with the mission of helping society reach a low- or zero-carbon future. They include Tongji University in Shanghai, China, which has been implementing building upgrades and energy-saving projects. In 2006, it saved about 10 million kilowatt-hours of electricity and reduced its carbon dioxide emissions by 9,200 tonnes, according to the university's vice-president Chen Xiaolong. And in 2008, he says, it installed a system to perform real-time monitoring of energy consumption in some 300 buildings across four campuses.

In southern Spain, Malaga University is installing solar panels that will produce a megawatt of energy to power the campus, along with geothermal energy and a trigeneration power plant to convert waste heat into power. The university aims to eventually meet all of its energy

needs through renewable energy, according to Rafael Morales, a university vice-rector and head of its sustainability programme.

In Britain, the University of the West of England in Bristol expects to have 100% of electricity on its academic sites coming from renewable sources by 1 October. From 2006 to 2007, the university cut its carbon emissions by 23%, says James Longhurst, an environmental scientist there. "We're on a journey," he says. "I don't think any of us are certain that we'll ever arrive, but we're on a journey towards being more sustainable."

In the United States, some of the most aggressive schools in the campus sustainability movement, such as Emory and Harvard, have chosen not to sign the presidents' climate commitment. In part, that's because many are sceptical of the commitment's focus on a zero-carbon goal. Reaching carbon neutrality will require schools to buy offsets, which are often criticized because they allow a polluter to pay a fee to support a green activity to 'offset' the polluter's carbon transgressions. "There's no way to become carbon neutral without buying offsets, mathematically," says Pearlman.

Buying offsets is still a fairly new and unregulated practice, so some are concerned that it could take the place of more meaningful emissions cuts. "Until it's better regulated, we didn't feel comfortable that we could say we knew exactly where every dollar of that was going," says Emory's Howett. But Cortese contends that over time, as schools make larger investments in green technologies and find better ways to cut carbon, fewer offsets will be necessary.

Model institutions

Many schools also see themselves as a test bed for green living from which communities and cities can learn. In Atlanta, a city notorious for traffic congestion and poor air quality, Emory is setting aside more than half of its campus as protected green space, working to create a bike culture, and providing incentives for its employees to ride buses powered by used cooking oil from its campus cafeterias. Harvard has developed a \$12-million revolving loan fund for sustainability projects, which doles out up to \$500,000 per project. Within just a few years, the work has saved nearly \$4 million annually and some 25,000 tonnes of carbon dioxide equivalent, says Henriksen. She says she has fielded calls from foundations and corporations and spoken to city managers who are thinking of setting up similar loan funds. And in 2007, Middlebury completed a renovation of its Franklin Environmental Center, housed in an 1870s farmhouse near the centre of the campus, as a model of sustainable design for those who want to go green while retaining the character of the region's architecture.

Institutions do not seem to be shying away from their commitments, despite the current financial downturn. Paul Fonteyn, president of Green Mountain College, says the school's new biomass-fuelled plant will save \$250,000–300,000 per year in heating costs. "I don't see how you can afford not to do this kind of activity," he says. Amy Johns, an environmental analyst at Williams College in Williamstown, Massachusetts, agrees. Although the financial belt-tightening has made some projects more challenging, she says, "a lot of them do have a pretty solid payback, so even in the hard financial times they can be pretty appealing."

Jack Byrne, director of the sustainability integration office at Middlebury, says that the recession is driving his school to find more efficient ways to accomplish its green goals. "The one thing that has been clear in all of this is that sustainability is a core value," he says. "We're just going to be looking for more effective and efficient ways to do it with fewer people." ■

Amanda Leigh Mascarelli

See Editorial, page 146.

Export-control laws worry academics

US researchers hope planned reforms will reduce the risk of prosecution.

Academics in the United States are hoping that pending legislation and a presidentially mandated review could provide long-sought relief from export laws they believe hamper international scientific cooperation and research.

The defence and aerospace industries have long struggled with the seemingly Byzantine nature of export-control regulations, as has NASA, which has sought exemptions to cover its work on the International Space Station. The recent sentencing of two US physicists to prison underscores how academics can also face penalties for failing to comply.

Many see the new administration of President Barack Obama as an opportunity to jump-start reforms. The Foreign Relations Authorization Act, passed by the House of Representatives in June, ordered a comprehensive assessment of the arms export-control system and would allow the president to remove satellites from the US munitions list — thus potentially easing life for the many academic space scientists who work on satellites. The bill has not yet passed the Senate, but separately on 13 August the White House announced that Obama was directing “a broad-based interagency process for reviewing the overall US export-control system”.

Previous efforts to change the laws have foundered amid congressional opposition, but reforms may go through this time, says Fred Tarantino, president of the Universities Space Research Association in Columbia, Maryland. White House reviews, such as the one on export control, typically take six to nine months to complete.

The export-control regime is split between two agencies: the commerce department, which is responsible for licensing dual-use items — those that have civilian and military applications — and the state department, which administers technologies deemed to be military items under the International Traffic in Arms Regulation (ITAR). For many academics, ITAR is the heart of the problem: divided into broad categories, it focuses on types of technologies rather than specific items, leaving many unsure as to what is covered.

Particularly confounding is the concept of a “deemed export”, which means the release of technical data or information to a foreign national, even if it takes place in the United States. Deemed exports thus potentially cover a professor lecturing to a class that includes



S. DOERING/VISUM/STILL PICTURES

Imparting military-sensitive information to foreign students can land professors in jail.

foreign nationals. “When you’re dealing with interns and graduate students, most people don’t understand how you can export an item by letting your graduate student know about it and look at it,” says Jim Barger of the law firm Frohsin and Barger, based in Birmingham, Alabama. “It’s counter-intuitive, and counter to academic discourse and exploration.”

Take, for example, the case of John Reece Roth, a physicist at the University of Tennessee, Knoxville, who was convicted this year of violating the Arms Export Control Act (AECA). At issue was Roth’s research on plasma actuators designed to reduce drag on unmanned aircraft. Roth, who was working for a university spin-off company under a US Air Force contract, allowed two graduate students — one from China and another from Iran — access to what the government determined was controlled technical data. He also took a laptop containing information about the work with him to China, where he was giving lectures on the topic. Roth was found guilty and sentenced in July to four years in federal prison. Daniel Max Sherman, a former colleague who cooperated with investigators, was sentenced in August to 14 months in prison.

“I feel this case is an anomaly,” says Robert Kovac, the state department’s managing director of the directorate of defence trade controls. “Increased outreach efforts, as well as the publicity associated with the case, have led to more awareness of AECA and ITAR requirements

within the university community.”

Indeed, university employees working on projects involving controlled technology should pay attention, warns Anupam Srivastava of the University of Georgia’s Center for International Trade and Security at Athens. “The penalty,” he says, “clearly suggests the federal government is serious about prosecuting cases.”

Concerns over prosecution have even led some academics to self-censor when teaching, particularly in the area of satellites, which have been under the control of the state department since 1999. That shift, which was prompted by a satellite manufacturer illegally sharing technical data with China about the failure of a Long March rocket, had an immediate effect on university work in the area. “There are things I was once comfortable talking about in class, and I’m not comfortable with anymore,” says Thomas Zurbuchen, a professor of space science and aerospace engineering at the University of Michigan in Ann Arbor.

As an example, Zurbuchen says he can talk in class about the physics of how radiation affects silicon in a circuit, but not how to solve that problem — because that would get into the specifics of manufacturing, something covered under ITAR. When it comes to the lab, those problems become even more complex; Zurbuchen says that he is forced to exclude foreign graduate students from working on space hardware altogether. “I’m liable myself,” he says, “not the university.”

Sharon Weinberger

Ethics scrutiny needed for Chinese–European projects

Biomedical research collaborations between Europe and China need greater ethical oversight to combat unregulated stem-cell therapies and prevent the exploitation of clinical-trial participants. That's the message from a group of bioethics experts who are part of the Chinese–European BIONET project, a partnership set up to examine scientific collaborations between the regions. Over the past three years, it has run a series of workshops in China to produce a set of best-practice guidelines for scientists working in fields such as reproductive and regenerative medicine, stem-cell research and human-tissue biobanking.

The group's draft recommendations, presented at the final BIONET meeting in London on 2–4 September, include a call for a joint advisory body made up of experts from participating countries, to offer advice and monitor research practices. The body could be financed by funding agencies, research institutions and state authorities, BIONET suggests.

"We have no police force," says BIONET member Ole Döring, an ethicist at the German Institute of Global and Area Studies in Hamburg. "We are proposing that if you install a body that would supervise and provide guidance, just the fact it exists will help create transparency."

The BIONET expert group warns that legal, political, social and cultural differences between European nations and China can lead to "multiple standards and even to gaps in between governance regimes". BIONET coordinator Nikolas Rose, a sociologist at the BIOS centre at the London School of Economics, says that there is a pressing need to address such issues because "the number of Chinese scientists who are collaborating with European scientists is growing at a massive rate". A 2006 study by the consultancy Evidence, based in Leeds, UK, shows that the number of publications co-authored by researchers in China and the European Union rose from 1,320 to 4,568 between 1996 and 2005.

But Rose insists that the BIONET recommendations are not an attempt to force China to adopt Western research standards. "China is not the 'Wild East',

it is not an ethics-free zone," he says.

The recommendations come less than a month after the China–UK Research Ethics (CURE) committee of the UK Medical Research Council (MRC) produced its own report on the subject, concluding that there is "comparatively little" inspection or review of compliance with research regulations in China. Qi Guoming, vice-chairman of the Chinese Medical Association and chairman of the medical ethics committee at the Chinese Ministry of Health, told the conference that the ministry was trying to come up with "more concrete regulations" for medical research, and that BIONET's recommendations could guide that process.

In May, for example, China toughened up its regulation of stem-cell therapies (see *Nature* 459, 146–147; 2009). But there are still more than 100 institutions in China that

continue to charge patients thousands of dollars for unproven stem-cell treatments, says Qiu Renzong, a bioethicist at the Chinese Academy of Social Sciences and co-chairman of the BIONET expert group.

BIONET's list of 30 recommendations includes establishing protocols to ensure that clinical trials of unproven therapies, such as stem-cell treatments, are not presented to patients as a cure. Research subjects should not be coerced into taking part in clinical trials, and all trial data should be published. BIONET also proposes that international ethical standards should be reflected by national regulation where possible, and that biobanks should ensure that any donors are fully informed about how their tissue will be used. The group adds that patients involved in clinical trials must have access to any beneficial therapies after the trials finish.

"Many of these recommendations reflect standards we would set for funding international collaboration," says Catherine Elliott, the MRC's head of clinical research support and ethics who coordinated the CURE report. "Some, however, would require much wider action and implementation than a single funder can provide." The new recommendations, she says, will trigger that wider discussion. ■

Daniel Cressey

"China is not the 'Wild East', it is not an ethics-free zone."

Toxicity testing gets a makeover

Europe aims to make chemical-exposure studies more predictive while using fewer animals.

ROME

The European Commission has revealed details of a major new research programme to develop a modern, high-throughput approach to repeat-dose toxicity testing.

Pressure to launch such an effort arose because the commission had drafted conflicting pieces of legislation, which demanded more extensive safety testing of chemicals while also requiring less use of animals in those tests. The programme, says the commission, will help to reconcile these goals.

"Faster, cheaper and more reliable alternative methods will contribute to increased safety" while reducing the use of animals, says a commission communiqué issued in Rome last week at the World Congress on Alternatives and Animal Use in the Life Sciences, where the €25-million (US\$36-million) programme was presented.

Two items of European legislation present particular dilemmas to industry. One is the 2006 Registration, Evaluation, Authorisation and Restriction of Chemical Substances (REACH) directive, which requires retrospective testing of chemicals that are being marketed, to a point that many think overburdens existing testing capacities (see *Nature* 460, 1065; 2009). The other is the 2003 amendment to the 1976 cosmetics directive, which phases out all testing of cosmetic ingredients on animals by 2013. The legislation also applies to imported products marketed in Europe.

Now, in the first agreement of its kind, industry will match the commission's funds through Colipa, the consortium of Europe's cosmetics, toiletry and perfumery industries based in Brussels. The total €50-million pot represents the largest-ever injection of money into the development of alternative toxicity testing.

The cosmetics industry is not particularly happy about coughing up the money when the chemicals industry is not doing the same. "Of course it is not fair," says one top representative of a cosmetics company, speaking on condition of anonymity. "But the legislation itself is not fair — the science is not there."

No one expects the new programme to be more than a modest start to the massive effort



By 2013, the European cosmetics industry will phase out animal testing.

"The programme puts toxicology on a new basis."

needed to rapidly and reliably test, with minimal animal use, for all possible adverse consequences of prolonged exposure to chemicals. "It will take 10 or 20 years before this is going to be translated," says meeting co-organizer Thomas Hartung, director of the Johns Hopkins University Center for Alternatives to Animal Testing in Baltimore, Maryland.

For instance, determining whether long-term exposure to a chemical causes cancer or neurological disease without using animals is much harder than the nearly completed work of replacing animals in single-exposure toxicity work. "You can't just go with a single endpoint — you have to know how the whole system works," says toxicologist Horst Spielmann of the Federal Institute for Risk Assessment in Berlin.

Advanced technology

The commission's call for projects intends to incorporate expertise in five areas not widely used in traditional toxicology. These include developing methods to reliably generate other types of human cells from stem cells, and developing cellular devices that simulate organs such as the heart, lungs or kidney. Other areas include systems biology and computational modelling.

Each area will be tackled by a single consortium of researchers. "We want to concentrate the money on the minimum number of labs who can do the work needed," says Jürgen Buesing, the commission official in charge of the programme.

Stem-cell researcher Jürgen Hescheler from the University of Cologne in Germany is one of those intending to apply for funding through the initiative. "The programme puts toxicology on a new basis and brings it into the right species: the human," he says.

A US initiative — the Tox21 programme coordinated by the Environmental Protection Agency and the National Institutes of Health — is also taking a high-throughput, systems approach to toxicology. With \$22 million for this year alone, it too aims to increase the predictive value of toxicity tests while

reducing animal use, and is prioritizing chemicals most in need of testing. "It is critical that Tox21, and data generated in other countries, are used in Europe so that there is no duplication," says Spielmann, who is running a project under Europe's seventh framework programme for research to ensure just that.

In the meantime, scientists at the Rome meeting said that steps must be taken now to reduce the unnecessary use of animals. Bennard van Ravenzwaay, head of toxicology at the German chemicals giant BASF in Ludwigshafen, says that tests should be abandoned if they add negligible predictive value to the battery of experiments already required by regulatory agencies. Such checks include the two-generation test for reproductive toxicology, in which the second generation uses many animals without providing useful information; the mouse cancer test, which provides negligible additional information beyond the rat cancer test; and developmental neurotoxicity checks.

Regulatory authorities can also engage in "intelligent toxicity testing strategies" to reduce the number of chemicals that need full testing, says Kees van Leeuwen of TNO, the Netherlands' applied research organization in Zeist. "We can reduce which chemicals may not need a full battery of testing, by optimizing the use of information from similar chemicals," he says.

Buesing says that national agencies and industry should be prepared to extend funding of alternative methods in toxicology in the near future. "Otherwise," he says, "our €50 million will have been wasted."

Alison Abbott

S. SANCHEZ/THE IMAGE BANK/GETTY

**EVIDENCE FOR MONOPOLES**

Materials with single points of north and south discovered.

www.nature.com/news

SCIENCEPHOTOS/ALAMY

World climate services framework agreed

GENEVA

A global framework to supply on-demand climate predictions to governments, businesses and individuals is moving closer to reality.

On 3 September, delegates representing 155 nations at the World Climate Conference in Geneva, Switzerland, agreed that a body should be established to supply such 'climate services' to users ranging from national governments to individual farmers. The service would be particularly helpful for developing nations, many of which lack access to the weather and climate observations needed to plan their strategies for adapting to climate change.

Over the next four months, an independent task force set up by the World Meteorological Organization will work out how to make this vision a reality. A 12-month consultation process with signatory nations will follow.

"It's about time we got serious," says

climatologist Jonathan Overpeck of the University of Arizona in Tucson. "We can save wealth and properties if we get climate information into the hands of decision-makers."

But a global climate service will face a host of scientific and political hurdles. Negotiating data collection and sharing among member states will be a big challenge, for example. Some countries are already baulking at the suggestion that they will need to supply the service with data, citing issues such as national security or commercial interests that would prevent disclosure. In response, Martin Visbeck of the Leibniz Institute of Marine Sciences at the University of Kiel in Germany says that one option would be to allow "data of convenience tailored for specific purposes [to] be commercialized", while allowing "fundamental information to be freely available".

Climate scientists will also have

to improve the quality of the climate projections that the service could provide. Today's global climate models predict how climate variables, such as temperature and rainfall, will change over the coming century at scales of several hundred kilometres. But scientists are hopeful that with further research they could bring that down to just tens of kilometres, covering timescales of a decade or less.

In the meantime, individual nations are forging ahead with their own climate-services centres. In July, Germany opened a centre in Hamburg, and the United States is discussing a national climate service. ■

Olive Heffernan

For a longer version of this story, see
<http://tinyurl.com/climate-service>.

Correction

The News Feature 'Last chance clinic' (*Nature* **460**, 1071-1075; 2009) inadvertently located Massachusetts General Hospital in Cambridge. It is in Boston.

Empty archives

Most researchers agree that open access to data is the scientific ideal, so what is stopping it happening? **Bryn Nelson** investigates why many researchers choose not to share.



In 2003, the University of Rochester in New York launched a digital archive designed to preserve and share dissertations, preprints, working papers, photographs, music scores — just about any kind of digital data the university's investigators could produce. Six months of research and marketing had convinced the university that a publicly accessible online archive would be well received. At the time of the launch, the university librarians were worried that a flood of uploaded data might swamp the available storage space.

Six years later, the US\$200,000 repository lies mostly empty.

Researchers had been very supportive of the archive idea, recalls Susan Gibbons, vice-provost and dean of the university's River Campus Libraries — especially as the alternative was to keep on scattering their data and dissertations across an ever-proliferating array of unintegrated computers and websites. "So we spent all this money, we spent all this time, we got the software up and running, and then we said, 'OK, here it is. We're ready. Give us your stuff,'" she says. "And that's where we hit the wall." When the time came, scientists couldn't find their data,

or didn't understand how to use the archive, or lamented that they just didn't have any more hours left in the day to spend on this business.

As Gibbons and anthropologist Nancy Fried Foster observed in their 2005 postmortem¹, "The phrase 'if you build it, they will come' does not yet apply to IRs [institutional repositories]."

A similar reality check has greeted other data-sharing efforts. Most researchers happily embrace the idea of sharing. It opens up observations to independent scrutiny, fosters new collaborations and encourages further discoveries in old data sets (see pages 168 and 171). But in practice those advantages often fail to outweigh researchers' concerns. What will keep work from being scooped, poached or misused? What rights will the scientists have to relinquish? Where will they get the hours and money to find and format everything?

Some communities have been quite open to sharing, and their repositories are bulging with

data. Physicists, mathematicians and computer scientists use arXiv.org, operated by Cornell University in Ithaca, New York; the International Council for Science's World Data System holds data for fields such as geophysics and biodiversity; and molecular biologists use the Protein Data Bank, GenBank and dozens of other sites. The astronomy community has the International Virtual Observatory Alliance, geo-

scientists and environmental researchers have Germany's Publishing Network for Geoscientific & Environmental Data (PANGAEA), and the Dryad repository recently launched in North Carolina for ecology and evolution research.

But those discipline-specific successes are the exception rather than the rule in science. All too many observations lie isolated and forgotten on personal hard drives and CDs, trapped by technical, legal and cultural barriers — a problem that open-data advocates are only just beginning to solve.

One of those advocates is Mark Parsons at

"We got the software up and running and said 'Give us your stuff'. That's when we hit the wall."

— Susan Gibbons

ILLUSTRATIONS BY J.H. VAN DER ENDONCK

the National Snow and Ice Data Center at the University of Colorado in Boulder. Parsons manages a global programme to preserve and organize the data produced by the International Polar Year (IPY) that ran from March 2007 to March 2009 and included an estimated 50,000 collaborators from more than 60 countries.

The IPY policy calls for data to be made available fully, freely, openly and on the shortest feasible timescale. “Part of what is driving that is the rapidness of change in the poles,” says Parsons. “If we’re going to wait five years for data to be released, the Arctic is going to be a completely different place.”

Reality bites

But reality is forcing a longer timescale. As soon as they began implementing the data policy, Parsons and his team encountered a staggering diversity of incoming information, as well as wide variations in the culture of data sharing. Fields such as atmospheric science and oceanography, Parsons says, have well-developed traditions of free and open access, and robust databases. But fields such as wildlife ecology and many of the social sciences do not. “What we discovered was that this infrastructure to share the data doesn’t really exist, so we need to start creating that,” Parsons says.

But his programme lacks the resources required to create that infrastructure on a large scale. So the team has resorted to preserving as much data as it can. It has delegated much of that job to national coordinators, or “data wranglers”, as Parsons calls them, who contact investigators and, “get the data branded and put in the IPY corral”.

One of the most successful data-wrangling countries has been Sweden, which formed a subcommittee to correct its early lag in collecting and then received national funding for its own IPY data archive. National coordinator Håkan Olsson, a specialist in remote sensing at the Swedish University of Agricultural Sciences in Umeå, says that the country’s archive is helping to house data from smaller, independent projects that would never reach large international databanks.

Nevertheless, he says, many Swedish researchers still don’t archive their data, or don’t put data in formats that make them easily searchable and retrievable. He faults the funding agencies too. “Unlike some other countries,” he says, “the research councils in Sweden do not yet have a practice to grant funds with the condition that data from the project is sent to a data centre.”

Even when wranglers can identify the data, it is not always obvious where the data should go. For example, says Parsons, “you would think that any snow and ice data would go into the

National Snow and Ice Data Centre”. But the centre’s funding is generally tied to specific data streams, he says, which means it can find itself in the position of accepting glacial data from a programme it has money for, while being forced to turn away similar glacial data from programmes where it does not.

Despite the launch earlier this year of the Paris-based Polar Information Commons to make polar data more accessible, Parsons says, that with all the “naïve assumptions”, the lack of planning and other unanticipated obstacles, properly managing the IPY data will require another decade of work.

In other fields, however, the main barriers to data sharing are concerns about quantity and quality. The US National Science Foundation’s (NSF’s) Laser Interferometer Gravitational-Wave Observatory (LIGO), for example, uses giant detectors in Louisiana and Washington to search for gravitational waves that might indicate the presence of rare phenomena such as colliding black holes or merging stars. LIGO is also working with the Virgo consortium, which operates a similar detector near Pisa, Italy.

Neither team has detected the signal they are looking for yet — but that’s not surprising: gravitational waves are expected to be extraordinarily faint. The key to detecting them is to eliminate every possible source of spurious vibration in the detectors, whether from seismic events, electrical storms, road traffic or even from the surf on distant beaches. It requires what Szabolcs Márka, a physicist at Columbia University in New York and the university’s lead scientist for LIGO, calls “a really paranoid monitoring of the environment”.

The question of what data should be shared has provoked strong debate within the LIGO and Virgo teams. Should they open up all their terabytes of data to outside scientists, including the torrents of environmental data? Or should they release just the cleaned-up data stream most likely to reveal a gravity wave? Would naïve outsiders fail to process the raw data adequately, leading to premature announcement of gravitational wave ‘discoveries’ that would hurt everyone’s credibility? Or would the extra eyes bring fresh perspective to the search?

“I’m torn,” says Márka, who says that the precise terms of data sharing are being negotiated with the project’s funders. “We don’t just have to analyse the data, we need to make sure the data are right.”

How data should be shared is also a substantial problem. A prime example is the issue of data standards: the conventions that spell out exactly how the digital information is

formatted, and exactly how the contextual information (metadata) is listed.

In some disciplines it is comparatively easy to agree on standards, says Clifford Lynch, executive director of the Coalition for Networked Information based in Washington DC, which represents academia on data and networking issues. “If you look at something like the sequencing of a genome, there’s a whole lot of tacit stuff that’s already settled,” he says. “Sequencing one genome is very similar to sequencing another.” But for other groups

— say, environmental scientists trying to understand the spread of a pollutant — the choice of common standards is far less obvious.

The all-too-frequent result is fragmented and often mutually incomprehensible scientific information. And that,

in turn, stifles innovation, says James Boyle, a law professor at Duke University in Durham, North Carolina, and a founding board member of Creative Commons, a non-profit organization that supports creative content sharing.

“We don’t just have to analyse the data, we need to make sure the data are right.”

— Szabolcs Márka

Always somebody smarter

“Researchers generally create their own formats because they believe that they know how their users want to use the data,” says Boyle. But there are roughly a billion people with Internet access, he says “and at least one of them has a smarter idea about what to do with your content than you do”. For example, web users are using applications such as Google Earth to plot the spread of pandemics² or to collect information on the effects of climate change. All that is needed, says Boyle, are common languages and formats for data.

Perhaps not surprisingly, data-sharing advocates say, the power to prod researchers towards openness and consistency rests largely with those who have always had the most clout in science: the funding agencies, which can demand data sharing in return for support; the scientific societies, which can establish it as a precedent; and the journals, which can make sharing a condition of publication.

The trick is to wield that power effectively. The NSF, for example, has funded groundbreaking research into digital archiving, search and networking technologies. But its data-sharing policies for standard research grants, for example, have come under fire for being scattered and ad hoc; they are often stipulated on a per-project basis. Gibbons says she is especially disappointed with a 2003 mandate by the US National Institutes of Health (NIH), which could have dramatically changed the culture of data sharing. The mandate does require a

data-sharing plan for any grant worth \$500,000 or more in direct annual costs or an explanation of why sharing isn't possible. But details about how to make the data available were so vague, says Gibbons, that researchers soon stopped paying attention, content to sit back until someone got in trouble for not playing by the rules.

Officials at the NIH Office of Extramural Research reply that the data-sharing policy's 'vagueness' is, in fact, flexibility, an attempt to avoid forcing every research programme into a one-size-fits-all straightjacket. They note that the policy also recognizes that there may be valid reasons for not sharing, including concerns about patient privacy and informed consent.

The chicken or the egg?

Nonetheless, until data sharing becomes a requirement for every grant, says Daniel Gardner, a physiologist and biophysicist at the Weill Medical College of Cornell University, "people aren't going to do it in as widespread of a way as we would like". Right now, he says, "you can't ask large numbers of people to do it, because it's a lot of work and because in many cases the databases don't exist for it. So there is kind of a chicken and egg problem here."

One solution would be for agencies to invest in the infrastructure necessary to meet their archiving requirements. That can be difficult to arrange, says Boyle. "Infrastructure is the thing that we always fail to fund because it's kind of everybody's problem, and therefore it's nobody's problem." Yet some agencies have been pioneers in this area. One often-cited example is the Wellcome Trust, the largest non-governmental UK funder of biomedical research. Since 1992, its Sanger Institute near Cambridge has been developing and housing some of the world's leading databases in genomics, proteomics and other areas.

Another prominent example is the NIH's National Library of Medicine, which in 1988 established the National Center for Biotechnology Information (NCBI) to manage its own collection of molecular biology databases, including the GenBank repository. James Ostell, chief of the NCBI's Information Engineering Branch, likes to show a colour-coded timeline of contributions to GenBank since its founding in 1982 — a progression that dramatizes the fast-evolving

history of genetic sequencing. Ostell points out thick waves of colours flowing from the left side of the chart. Representing traditional sequence divisions such as viruses, rodents, primates, plants and bacteria, they dominated GenBank's contents for years. Other sequences, produced by faster techniques, began to put in appearances in the mid 1990s. Then in late 2001 a sudden surge of green, representing DNA snippets derived from whole-genome shotgun sequencing, quickly took over. By 2006, the green accounted for more than half of the database's contents.

Keeping up with ever-shifting technology has created its own set of challenges, says Ostell. "Nobody has infinite resources. And storing electronic information over time is a dynamic process. If you try to look at a file that you wrote with a word processor 20 years ago, good luck." In the same way, if a data set isn't readable by the latest version of a database, it isn't usable. So an archive may well have to choose between tossing old data out, and paying to preserve the out-of-date software required to make sense of them.

Even more challenging are the legal minefields surrounding personal data and privacy. The need to protect human subjects has led to starkly different approaches. Some projects

openly share data, whereas others require researchers to navigate a labyrinthine approval process before granting access. The NCBI has tried to build such requirements into its newer databases. A case in point is its database of Genotype and Phenotype (dbGaP), which archives and distributes the results of genome-wide association studies, medical DNA sequencing, molecular diagnostic assays and almost any-

thing else that relates people's traits and behaviours to their genetic makeup. The dbGaP allows open access to summaries and other forms of information that have been stripped of personal identifiers. But it grants controlled access

to personal health information only after a researcher has been approved by a formal review committee.

Novel meaning

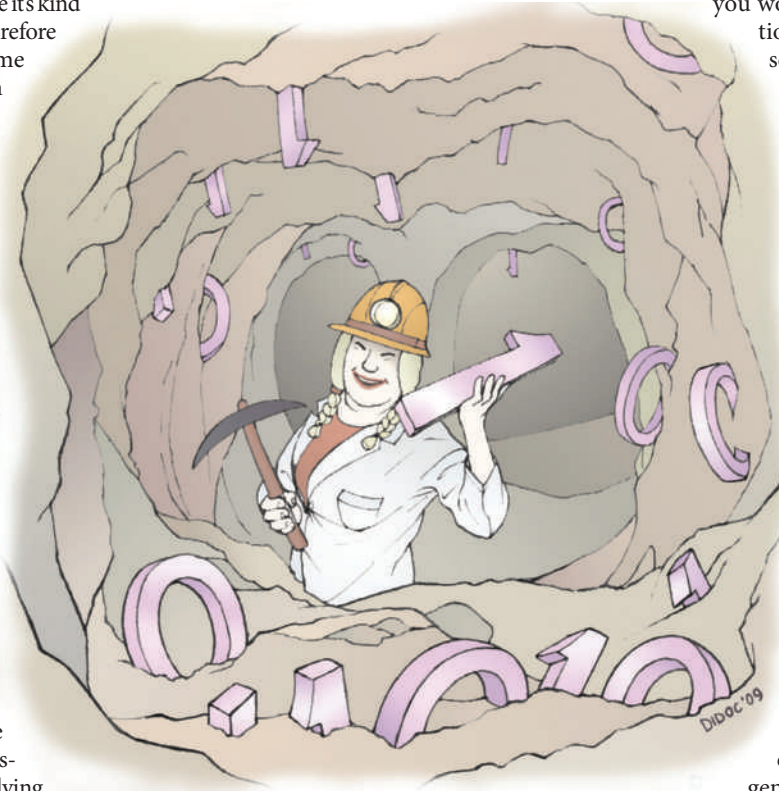
Such measures can be cumbersome, says Ostell. Yet the benefits of sharing far outweigh the costs. Some of GenBank's early sequences, for example, included genes from yeast and *Escherichia coli* labelled as DNA repair enzymes. Years later, researchers studying human colon cancer made a link between mutations in patients and those same enzymes³. "If you just did a literature search, you would never make that connection," Ostell says. "But when you search on the basis of their genes, suddenly you connect meaning in a way that's novel, which is the basis of discovery."

Sharing is obviously easier when the expectations are clear, and many scientists point to a 1996 meeting in Bermuda as a defining moment for genomics. At the meeting, leaders working on the Human Genome Project hammered out a set of agreements known as the Bermuda principles. Chief among them was the stipulation that sequences longer than 1,000 base pairs be made publicly available, preferably within 24 hours.

The Bermuda principles, in turn, built on the foundations laid a decade earlier by the editors of journals such as *Nucleic Acids Research*, who spurred the early development of GenBank and other genomic repositories by requiring

"At least one of the people out there has a smarter idea about what to do with your content than you do."

— James Boyle



researchers to deposit their data there as a precondition for publishing. Newer journals, such as the open-access Public Library of Science journals, have made publication contingent on making the data “freely available without restriction, provided that appropriate attribution is given and that suitable mechanisms exist for sharing the data used in a manuscript”. The journal *Neuroinformatics* devoted its September 2008 issue to data sharing through the NIH Neuroscience Information Framework. *Ecological Archives* publishes appendices, supplements and data — related to studies appearing in other ecology journals — which include the metadata needed to interpret them. (*Nature* journals require authors “to make materials, data and associated protocols promptly available to readers without preconditions”.)

Yet the journals’ power to compel data sharing and scientific culture change is not absolute. In March 2009, for example, the journal *Epidemiology* felt able to call only for a “small step” towards more openness. “We invite our authors to share their data and computer code when the burden is minimal,” said an editorial⁴ in that issue.

“We believe that data sharing is a matter of time,” says Miguel Hernán, an epidemiologist at Harvard University and a co-author of the editorial. But prematurely forcing a sharing requirement on authors “would be suicidal”, he warns, especially with unresolved concerns over patient confidentiality. They would simply submit their papers somewhere else.

Another issue facing journals and data banks is how to ensure proper citations for data sets. “The one thing that people clearly care about in the sciences is attribution,” says Boyle. Without an agreed-on way of assigning credit for original data falling beyond the parameters of a publication, however, it’s no wonder that scientists are reluctant to share: their hard work may never be recognized by their employers or by granting agencies. Worse yet, it could be poached or scooped.

This is one place that technology might help, says Boyle. He points to a music site associated with Creative Commons known as ccMixter, in which users can upload an a cappella chorus, a bass line, a trumpet solo or other musical samples. Users are free to remix the samples into new tracks. But when they do, the program automatically keeps a continuous credit record.



So why not implement a similar system that would add a link back to a database every time a researcher repurposed some data? It wouldn’t necessarily solve the problem of scooping, Boyle says, “but it aligns the social incentives with the individual incentives”. It could also provide a feasible way for universities or funding agencies to track the value of a researcher’s data.

International agreement

Other Creative Commons tools are already making their way into international scientific agreements. In May, for example, Creative Commons’ CC0 licence was endorsed by participants at a meeting in Rome on resource and data sharing within the mouse functional genomics community. The licence, which allows its users to “waive all copyrights and related or neighbouring rights” and thereby share more of their work, has been translated into dozens of languages.

As welcome as such developments are, however, Boyle points out that the creation of the legal and technical infrastructure to accommodate researchers’ data-sharing concerns is a huge task, and should not be left solely to non-profit organizations and individual universities. Nor should it be left to the funding agencies’ grant-by-grant allocations for data sharing. It will require major government investments, starting with demonstration projects to explore how sharing can best be done. “What we need is a working example that you can point to,” he says.

If William Michener has his way, a virtual data centre funded by the NSF and hosted by his university will be one of those examples. DataONE (Data Observation Network for Earth) exists only on paper, but a five-year, \$20-million grant through the NSF’s DataNet programme will help to turn it into an

open-access database focusing on biology, ecology and environmental science data. Four other \$20-million archives are planned under DataNet’s first phase.

Michener, director of e-science initiatives for University Libraries at the University of New Mexico, Albuquerque, and a leader of DataONE, says that the archive is designed to accommodate many of the orphan data sets that have yet to find a home, and will target resource-strapped colleges, field stations, and individual or small teams of scientists.

In the longer term, the DataONE consortium, which encompasses two dozen partner institutions in the United States, the United Kingdom, South Africa, Australia and Taiwan, will explore business models that could sustain the archive well beyond its initial grant and potential five-year renewal. Among the plans under consideration are a fee-for-service set up, a membership requirement for participating entities and the solicitation of external grants for education and outreach.

DataONE’s success, however, may depend on overcoming the same ambivalence among researchers that has bedevilled the University of Rochester and other builders of public databases. Although a strategy is still being worked out, Michener envisions a combination of workshops, seminars, websites and other educational tools to help clarify the how and why of sharing. But one archive can only do so much. Larger efforts will be required to tackle what Michener sees as the overriding challenge: “Changing the culture of science from one

“We need to change the culture of science to one that equally values publications and data.”

—William Michener

where publications were viewed as the primary product of the scientific enterprise to one that also equally values data.”

Without that cultural shift, says Gibbons, many digital archives are likely to remain little more than stacks of empty shelves.

Bryn Nelson is a freelance science and medical writer based in Seattle, Washington.

1. Foster, N. F. & Gibbons, G. *D-Lib Magazine* doi:10.1045/january2005-foster (2005).
2. www.nature.com/avianflu/google-earth/index.html
3. Marra, G. & Boland, C. R. *Gastroenterol. Clin. North Am.* **25**, 755–772 (1996).
4. Hernán, M. A. & Wilcox, A. J. *Epidemiology* **20**, 167–168 (2009).

See Opinion, pages 168 and 171, and online special at: <http://tinyurl.com/dataspecial>.



MOUTH TO MOUTH

Hagfish and lampreys are the only surviving fish without jaws. And they could solve an evolutionary mystery, finds **Henry Nicholls**.

In the basement of the National Museum of Natural History in Paris, two men come to a standstill in the long, gloomy corridor nicknamed 'the submarine'. Philippe Janvier, a senior palaeontologist at the museum, unlocks a door, flicks on the light and leads the way into the 'salle poissons', the room that houses the museum's impressive collection of fossil fish. His visitor, Shigeru Kuratani, is a developmental biologist at Okayama University, Japan, who usually studies the lamprey — one of only two groups of jawless fish with living members. But today, he has come to see some of its long-extinct cousins.

As Kuratani peers at the vivid impression of a jawless fish etched into rock around 400 million years ago, the two get talking. Janvier suggests that Kuratani try to get his hands on an embryo from a hagfish, the only other group of jawless fish that still survives. Few researchers have been able to do it; if Kuratani could, it might resolve a taxonomic dispute that has troubled scientists for more than a century.

For several years after that encounter in 2000, Kuratani mulled it over. Then, in 2004, he took on Kinya Ota as a postdoc at his lab at the RIKEN Center for Developmental Biology in Kobe, and set him the task of succeeding where dozens had failed. "If you get embryos,"

Kuratani assured him, "just one or two, it will make a very important paper."

Kuratani and Janvier are not alone in their obsession with hagfish and lampreys. To a dedicated group of biologists these 'living fossils' are highly prized for what they promise to reveal about some of the earliest events in vertebrate evolution. And advances in developmental biology and molecular genetics are starting to fulfil that promise.

Hagfish and lampreys take researchers back around 500 million years to a time when the first jawed vertebrates, or gnathostomes,

evolved along with a truly 'vertebrate' body plan. The gnathostomes eventually dominated; apart from the hagfish and lampreys, the jawless 'agnathans' went extinct. The question is how exactly the split occurred between the hagfish, lampreys

and gnathostomes (pictured above, left to right), and the conflict between researchers' answers has been described as "one of the most vexing problems in vertebrate phylogenetics"¹. "We are struggling with this discrepancy at the very base of the vertebrate tree and we can't get out of it right now," says Janvier. "We have to find more and different kinds of data."

It is a problem with a history. In 1806, French zoologist André Duméril decided that the striking but similar mouthparts of hagfish and

lampreys meant that they should be grouped together (see 'Two trees') and called cyclostomi, or 'round mouths'. But from the 1970s onwards, morphologists began to have their doubts. Looking beyond the mouth, they found that adult lampreys boast a suite of characteristics that hagfish don't have, including elements of a vertebral column, an ability to control water content by osmoregulation, and the presence of true lymphocytes, a type of white blood cell. This suggested a tree in which lampreys were more closely related to gnathostomes than to the more primitive hagfish lineage.

That might have been the end of it, were it not for molecular biology. From the first trickle of sequence data to today's bioinformatics deluge, just about every molecular analysis suggests that Duméril was right after all: hagfish and lampreys are more closely related to each other than either is to gnathostomes. In this case, the last common ancestor of the two had a vertebral column and other characteristics, and these were secondarily lost by hagfish.

Only one of these trees can be right. It is rather important which one, as the precise route that these branches took has a profound effect on what can be inferred about the evolution of early vertebrates. For many researchers, the morphologists' tree is rather more alluring, as it would allow them to map out the events on the evolutionary path from headless invertebrates through hagfish with heads

"We are struggling with this discrepancy at the very base of the vertebrate tree."

— Philippe Janvier

K. Ota; M. Roggo/NaturePL.com; Stephen Frink Collection/Alamy

but no vertebrae, to lampreys with vertebrae but no jaws, to jawed gnathostomes (see 'Fossil finds'). But morphologists and molecular biologists — each of whom are staking out their own arrangements — seem unlikely to come to any kind of consensus. To Janvier, the idea of plugging these different types of data into a combined analysis doesn't make much sense.

A study earlier this year did combine them, and in doing so it illustrated the depth of the divide. Thomas Near, a molecular systematist at Yale University, was the first person to force morphological and molecular data sets into a single analysis¹. With molecular data pulled together from 4,638 ribosomal RNA sites and more than 10,000 amino acids, hagfish and lampreys emerge as undisputed sister groups. But the addition of just 115 morphological characteristics (from the skeleton and from the sensory, nervous and circulatory systems, for example) re-roots the tree, suggesting instead that lampreys are more closely related to gnathostomes. Near says that it is probably the molecular data that are giving the misleading result, because of difficulties in using DNA and protein sequences to shed light on events that occurred over a very short timescale — hagfish, lampreys and gnathostomes all diverged within a few million years — relative to the hundreds of millions of years that have passed since then. The findings give reason, the paper concludes, "to view the strong support for cyclostome monophyly inferred from molecular data sets with a measured degree of skepticism"¹. So how to resolve the problem?

Start at the beginning

That's where Kuratani's embryos come in. One way of working out evolutionary relationships is to look for a common developmental trajectory in the shape and growth of embryos — a field called 'evo-devo'. "As a general rule there is a danger of looking at an adult and assuming homology between different structures," Kuratani says. "Embryology cuts through that problem."

What researchers want to do is line up the embryos of hagfish, lampreys and a descendant of an early jawed vertebrate — such as the tropical brown-banded bamboo shark

Fossil finds

Palaeontologists have been trying to resolve early events in vertebrate evolution by attempting to illuminate the journey from a jawless to a jawed existence. A pressing challenge has been to make sense of the masses of extinct jawed fish that seem to shoal around this evolutionary transition and place the acquisition of anatomical features onto a timeline. And some fossil finds raise questions about the validity of groups such as the heavily armoured placoderms and the spiny-shark-like acanthodians.

The discovery earlier this year of *Guiyu oneiros* (pictured), an ancient example of a bony fish, has certainly shaken things up⁵. The appearance of this really complex fish around 419 million years ago suggests that most major events in the evolution of modern vertebrates occurred much earlier than was thought. The remains of cartilaginous, lobe-finned and ray-finned fish should populate the fossil



record prior to *G. oneiros*, but nobody has found them.

Where are they? A very real possibility is that some of their remains — isolated teeth, scales and spines — have been unearthed but classified wrongly, says Michael Coates, a palaeontologist at the University of Chicago in Illinois. "Instead of discovering early sharks, we identify these fragments as placoderms or acanthodians," he says. "We miss the opportunity to track the early evidence of how sharks diverged from bony fishes."

A recent report lends support to this view. Earlier this year, Martin Brazeau, then a PhD student at Uppsala University in Sweden,

published his analysis of an overlooked acanthodian braincase that boasted a combination of characteristics of several different groups⁶. Brazeau's work suggests that neither placoderms nor acanthodians are bona fide biological entities, bringing the groups one step closer to disintegration, says Coates. An artificial grouping could be an exciting opportunity for palaeontologists. These miscellaneous groups could contain a wealth of information that will resolve the order and timing at which key characteristics such as jaws, teeth, paired fins and internal fertilization were acquired during early vertebrate evolution. **H.N.**

(*Chiloscyllium punctatum*) — and compare not only their morphological development but also their patterns of gene expression. But getting hold of embryos from hagfish, lampreys or a species representative of early gnathostomes has proven extremely tricky.

For many years, lampreys have been the only cyclostome that evo-devo biologists have had to work with. These slender animals spend most of their lives as mud-dwelling, filter-feeding larvae before metamorphosing into toothy adults that often latch onto fish, rasping them with their tongue until they make enough of a wound to suck blood. The embryos are available for only a few weeks a year, so are difficult to obtain. For several years, members of Marianne

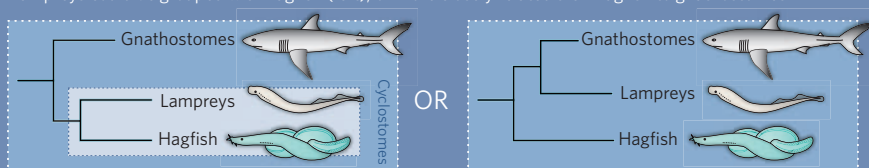
Bronner-Fraser's lab at the California Institute of Technology in Pasadena, for example, collected adults in the field, massaged the gametes from them, then performed *in vitro* fertilization and rudimentary investigations of lamprey development on the spot. Then, Bronner-Fraser says, "we realized the adults could be FedExed", and have since worked out how to extend their reproductive period in the lab.

Hagfish embryos have been even more challenging. The natural habitat of the few dozen described species is in the sludge at the bottom of the ocean. So elusive are hagfish that in the 1860s, the Danish Royal Academy of Sciences and Letters in Copenhagen offered a reward for the first person to work out the reproductive and developmental secrets of the Atlantic hagfish (*Myxine glutinosa*). Almost a century and a half later, the prize is still unclaimed.

After Ota accepted Kuratani's challenge, his first stop was the local fishermen. One of them agreed to supply some adult Japanese inshore hagfish (*Eptatretus burgeri*). Ota put them in a large tank back at Kuratani's laboratory, placed oyster shells and plastic drainpipes in the bottom to give the hagfish somewhere to hide, then regularly hauled the hideaway out on a

TWO TREES

Lampreys could be grouped with hagfish (left), or more closely related than hagfish to gnathostomes.



rope to check for eggs. Finally, Ota found what he was looking for: a cluster of eggs deposited on the fine-grained sand². A year later, the embryos became visible; a *Nature* paper followed soon after³.

The researchers did not resolve the phylogenetic debate, though. The paper showed that in hagfish, development of the embryonic structure called the neural crest and expression of the genes there are very similar to what is seen in both lampreys and jawed vertebrates. Since then, further embryos have been forthcoming. "We are trying to identify the basic design of vertebrates," says Kuratani. "If we can resolve this phylogenetic relationship between lamprey, hagfish and shark, then we can nail what kind of shape would have been there in the latest common ancestor of vertebrates," he says.

Head to head

For now, he and Ota are concentrating on comparing the heads of lampreys and hagfish. The head is a highly specialized structure that "defines the vertebrates", Kuratani says, because building features such as nostrils and a mouth opening required specific and "elaborate" developmental changes during evolutionary history. The researchers are comparing the first pharyngeal arch, for example — a nub of tissue that appears early in the life of vertebrate embryos and gives rise to the jaw and other head structures. This could show whether, as they suspect, the patterns

of gene expression seen in the developing lamprey more closely resemble those observed in gnathostomes.

While some researchers focus on embryos, others are concentrating on genetic sequences. With genome sequencing for the hagfish pencilled in by the US National Human Genome Research Institute in

Bethesda, Maryland, the sea lamprey already sequenced to 6× coverage and a draft genome assembled for the elephant shark (a jawed reference point), there is already a mass of genetic evidence to bring to the problem.

But as Near found in his analysis, standard sequence data may not be enough. So some researchers are now looking to other molecular data, in particular micro RNAs (miRNAs) — the snippets of RNA that are not translated into proteins but perform important regulatory functions. miRNAs are continually added to the genomes of complex eukaryotes such as vertebrates and, once they find a use in a genetic network, they are highly conserved by evolution and rarely lost. This means that if researchers can identify which miRNAs are present — much as a morphologist would score the presence or absence of a physical characteristic — they can potentially reveal more about when the two lineages split than they can by comparing in detail

Lamprey embryos (green) and hagfish (eggs shown, brown) could reveal similarities in development.

other genetic sequences, which requires complex statistics. "There's no other set of molecular data like it," says Kevin Peterson, a palaeobiologist at Dartmouth College in Hanover, New Hampshire. "Unlike other molecular data, it's treated as a set of binary characters," he says. "The morphologists can deal with these data."

A couple of years ago, Peterson compared the miRNA sequences of numerous organisms, including invertebrates such as sea urchins, and vertebrates such as sharks. He unearthed an extraordinary pulse of miRNA acquisition somewhere between 550 million and 505 million years ago — at around the same time that complex vertebrate features such as the head, gills, kidneys and thymus evolved⁴. "Something really amazing was happening to the vertebrate genome at that time," says Peterson. He says that acquisition of these miRNAs could have allowed cells to adopt more complex regulatory systems and to develop new and diverse cell functions. "It's those miRNAs that I would argue allow you to get novel cell types," he says.

But can this help solve the hagfish–lamprey problem? Peterson has been working with palaeobiologist Philip Donoghue of Bristol University, UK, to produce a library of the miRNAs present in hagfish, lampreys and some living gnathostomes — elephant shark, zebrafish and human. "We can use their presence or absence to finally resolve after 150 years or so the relationships between hagfish, lampreys and gnathostomes to work out the pattern of assembly of the body plan of jawed vertebrates," says Donoghue. The libraries have been sequenced and analysed, although neither Peterson nor Donoghue is giving away the result — yet.

On that cliffhanger, the story now rests. Whichever phylogenetic tree Peterson's results favour, he is hoping that it will be something that morphologists and molecular biologists can mull over together. "Our data clearly indicate that one answer is right," teases Peterson. "They unequivocally resolve the debate."

Henry Nicholls is a freelance writer based in London.

1. Near, T. J. *J. Exp. Zool.* doi:10.1002/jez.b.21293 (2009).
2. Ota, K. G. & Kuratani, S. *Zool. Sci.* **23**, 403–418 (2006).
3. Ota, K. G., Kuraku, S. & Kuratani, S. *Nature* **446**, 672–675 (2007).
4. Heimberg, A. M., Sempere, L. F., Moy, V. N., Donoghue, P. C. & Peterson, K. J. *Proc. Natl Acad. Sci. USA* **105**, 2946–2950 (2008).
5. Zhu, M. *et al.* *Nature* **458**, 469–474 (2009).
6. Brazeau, M. D. *Nature* **457**, 305–308 (2009).



In pursuit of hagfish embryos, Kinya Ota (front) approached local fishermen to obtain adult fish.

CORRESPONDENCE

Choking on carbon emissions from Greek academic paperwork

SIR—Selection processes for academic jobs are notoriously open to criticism, but in Greece they have the additional drawback of leaving a hefty carbon footprint.

Typically, selection committees for research institutes require applicants for a senior post to submit 11 paper copies of each of their publications (the Greeks' expansive view of publication sometimes includes texts of oral presentations) as well as of their birth certificate, national identity card (both sides), transcripts, translations of foreign degrees, and military and police reports.

In one recent case, the committee stopped the process to ask the minister of development to decide whether one worthy candidate should be excluded on the grounds that he had submitted his 66 publications on 11 CDs instead of on paper.

What is the environmental impact of this nonsense? If candidates have an average of 50 publications each, a single copy of these, plus the additional paperwork required, can add up to a package of 1,000 pages per candidate. Making 11 copies of each, in an election with, say, four short-listed candidates, generates 44,000 sheets of paper. Excluding the cardboard boxes necessary to transport them, this works out to some 378 kg of CO₂ per election (*Solid Waste Management and Greenhouse Gases* US Environmental Protection Agency, 1998). Almost all are later dumped in their original packaging, unopened.

University faculty positions are the worst, often being advertised at multiple levels. Candidates have to submit an identical package for each application level, copied to each voting member of the department. This can run to more than 100 complete sets of materials per candidate per position, contributing some 700 tonnes of greenhouse-gas emissions annually. Some of these

cases end up being decided by the courts, so the pollution escalates.

It would help if the European Union would step in to curtail such wasteful and irresponsible practices.

Costas Synolakis Viterbi School of Engineering, University of Southern California, Los Angeles, California 90089-2531, USA

e-mail: costas@usc.edu

Spyros Foteinis Laboratory of Natural Hazards, Technical University of Crete, Chania, Greece

Evolution pioneers: celebrating Lamarck at 200, Darwin 215

SIR—I take issue with the contention that Erasmus Darwin, the grandfather of Charles, tackled evolution only in poetic terms, as implied by Dan Graur and colleagues in their insightful Book Review ('In retrospect: Lamarck's treatise at 200' *Nature* **460**, 688–689; 2009).

Erasmus Darwin's most important contributions to evolutionary thought will be found in the very unpoetic prose of the first volume of his major medical and zoological treatise, *Zoonomia*, published in 1794.

Here, notably in Section 39, are discussions of deep time and the descent of all life from a single ancestor, bauplan homology among vertebrates, the analogy of artificial selection as a means of understanding descent with modification, and a brief but clear enunciation of the process of sexual selection.

One need only to look at the backlash against Erasmus Darwin's evolutionary ideas, in the savage political cartoons of James Gillray in 1798 and of others, to understand that — years before Lamarck made his contributions to evolutionary thought — Erasmus Darwin was triggering strong reactions for promoting a transformist view of biodiversity.

This year is justly celebrating the history-altering contributions of Charles Darwin. But it is equally

important to take stock of the critical intellectual steps before 1859 that made scientific and social acceptance of evolution possible.

Besides Erasmus Darwin and Jean Baptiste Lamarck, a host of other influential evolutionists, including Etienne Geoffroy Saint-Hilaire, Robert Chambers, Baden Powell, Herbert Spencer and Alfred Russel Wallace, deserve to be recognized (as well as read) for having laid a path to a modern view of descent with modification.

William E. Friedman Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado 80309, USA
e-mail: ned@colorado.edu

Evolution pioneers: Lamarck's reputation saved by his zoology

SIR—Work by Lamarck scholars over the past 20 years calls into question some of the assertions made by Dan Graur and his colleagues in their Book Review (*Nature* **460**, 688–689; 2009).

For example, far from being universally scorned, Jean Baptiste Lamarck became known as 'the French Linnaeus' during the 1820s. Speaking at Lamarck's funeral in December 1829, Etienne Geoffroy Saint-Hilaire remarked that the last years of the old naturalist's life had been brightened by the awareness of how much his work was appreciated in Europe, and especially in France (see www.lamarck.cnrs.fr).

During the 1820s, scientific, medical and cultural magazines discussed Lamarck's work at length. Even conservative commentators, who disliked Lamarck's veiled atheism, acknowledged his eminence as the foremost invertebrate zoologist of Europe. In Britain, several naturalists — including Darwin's first scientific mentor in Edinburgh, Robert Edmond Grant — bought Lamarck's works

and commented favourably on them. Lamarck's *Natural History of Invertebrates* (1815–22) became compulsory reading for hundreds of practitioners of the newly fashionable science, geology.

Furthermore, Lamarck can scarcely be said to be a deist, as your authors seem to argue. He did not deny that people had an idea of God, but as the only possible knowledge open to humankind was based on material substances and properties, nothing at all could be said of God. To Lamarck, nature had no purpose, no finality — in short, it was going nowhere.

Pietro Corsi University of Oxford, History Faculty, The Old Boys High School, George Street, Oxford OX1 2RL, UK
e-mail: pietro.corsi@history.ox.ac.uk

Religious belief and the history of science

SIR—I am concerned that the survey responses expressed in Gene Russo's Prospects article 'Balancing belief and bioscience' are irrelevant to gauging the influence of religion on the development of scientists (*Nature* **460**, 654; 2009).

Many of the great scientists renowned for developing entire scientific fields or theories were religious. For example, Gregor Mendel was a priest and Isaac Newton apparently spent as much time in religious contemplation as he did on calculus and physics. And Albert Einstein said: "Science without religion is lame, religion without science is blind."

As the works of most scientists today are not comparable with those of such luminaries, we should be cautious about using statistics on religious preference in judging scientific merit.

Scott Goode Department of Pathology, Baylor College of Medicine, S210 One Baylor Plaza, Houston, Texas 77030, USA
e-mail: sgoode@bcm.tmc.edu

OPINION

Prepublication data sharing

Rapid release of prepublication data has served the field of genomics well. Attendees at a workshop in Toronto recommend extending the practice to other biological data sets.

Open discussion of ideas and full disclosure of supporting facts are the bedrock for scientific discourse and new developments. Traditionally, published papers combine the salient ideas and the supporting facts in a single discrete 'package'. With the advent of methods for large-scale and high-throughput data analyses, the generation and transmission of the underlying facts are often replaced by an electronic process that involves sending information to and from scientific databases. For such data-intensive projects, the standard requirement is that all relevant data must be made available on a publicly accessible website at the time of a paper's publication¹.

One of the lessons from the Human Genome Project (HGP) was the recognition that making data broadly available before publication can be profoundly valuable to the scientific enterprise and lead to public benefits. This is particularly the case when there is a community of scientists that can productively use the data quickly — beyond what the data producers could do themselves in a similar time period, and sometimes for scientific purposes outside the original goals of the project.

The principles for rapid release of genome-sequence data from the HGP were formulated at a meeting held in Bermuda in 1996; these were then implemented by several funding agencies. In exchange for 'early release' of their data, the international sequencing centres retained the right to be the first to describe and analyse their complete data sets in peer-reviewed publications. The draft human genome sequence² was the highest profile data set rapidly released before publication, with sequence assemblies greater than 1,000 base pairs usually released within 24 hours of generation. This experience demonstrated that the broad and early availability of sequence data greatly benefited life sciences research by leading to many new insights and discoveries², including new information on 30 disease genes published prior to the draft sequence.

At a time when advances in DNA sequencing technologies mean that many more laboratories can produce massive data sets, and when an ever-growing number of fields (beyond genome sequencing) are grappling with their own data-sharing policies, a Data Release Workshop was convened in Toronto in May 2009 by Genome Canada and other funding agencies. The meeting brought together a diverse and multinational

group of scientists, ethicists, lawyers, journal editors and funding representatives. The goal was to reaffirm and refine, where needed, the policies related to the early release of genomic data, and to extend, if possible, similar data-release policies to other types of large biological data sets — whether from proteomics, biobanking or metabolite research.

Building on the past

By design, the Toronto meeting continued policy discussions from previous meetings, in particular the Bermuda meetings (1996, 1997 and 1998)^{3–5} and the 2003 Fort Lauderdale meeting, which recommended that rapid prepublication release be applied to other data sets whose primary utility was a resource for the scientific community, and also established the responsibilities of the resource producers, resource users, and the funding agencies⁶. A similar 2008 Amsterdam meeting extended the principle of rapid data release to proteomics data⁷. Although the recommendations of these earlier meetings can apply to many genomics and proteomics projects, many

outside the major sequencing centres and funding agencies remain unaware of the details of these policies, and so one goal of the Toronto meeting was to reaffirm the existing principles for early data release with a wider group of stakeholders.

In Toronto, attendees endorsed the value of rapid prepublication data release for large reference data sets in biology and medicine that have broad utility and agreed that prepublication data release should go beyond genomics and proteomics studies to other data sets — including chemical structure, metabolomic and RNA interference data sets, and to annotated clinical resources (cohorts, tissue banks and case-control studies). In each of these domains, there are diverse data types and study designs, ranging from the large-scale 'community resource projects' first identified at Fort Lauderdale (for which meeting participants endorsed prepublication data release) to investigator-led hypothesis-testing projects (for which the minimum standard should be the release of generated data at the time of publication).

Several issues discussed at previous data-

EXAMPLES OF PREPUBLICATION DATA-RELEASE GUIDELINES

Project type	Prepublication data release recommended	Prepublication data release optional
Genome sequencing	Whole-genome or mRNA sequence(s) of a reference organism or tissue	Sequences from a few loci for cross-species comparisons in a limited number of samples
Polymorphism discovery	Catalogue of variants from genomic and/or transcriptomic samples in one or more populations	Variants in a gene, a gene family or a genomic region in selected pedigrees or populations
Genetic association studies	Genomewide association analysis of thousands of samples	Genotyping of selected gene candidates
Somatic mutation discovery	Catalogue of somatic mutations in exomes or genomes of tumour and non-tumour samples	Somatic mutations of a specific locus or limited set of genomic regions
Microbiome studies	Whole-genome sequence of microbial communities in different environments	Sequencing of target locus in a limited number of microbiome samples
RNA profiling	Whole-genome expression profiles from a large panel of reference samples	Whole-genome expression profiles of a perturbed biological system(s)
Proteomic studies	Mass spectrometry data sets from large panels of normal and disease tissues	Mass spectrometry data sets from a well-defined and limited set of tissues
Metabolomic studies	Catalogue of metabolites in one or more tissues of an organism	Analyses of metabolites induced in a perturbed biological system(s)
RNAi or chemical library screen	Large-scale screen of a cell line or organism analysed for standard phenotypes	Focused screens used to validate a hypothetical gene network
3D-structure elucidation	Large-scale cataloguing of 3D structures of proteins or compounds	3D structure of a synthetic protein or compound elucidated in the context of a focused project

The Toronto statement**Rapid prepublication data**

release should be encouraged for projects with the following attributes:

- Large scale (requiring significant resources over time)
- Broad utility
- Creating reference data sets
- Associated with community buy-in

Funding agencies should facilitate the specification of data-release policies for relevant projects by:

- Explicitly informing applicants of data-release requirements, especially mandatory prepublication data release
- Ensuring that evaluation of data release plans is part of the peer-review process
- Proactively establishing analysis plans and timelines for projects releasing data prepublication
- Fostering investigator-initiated prepublication data release
- Helping to develop appropriate consent, security, access and

governance mechanisms that protect research participants while encouraging prepublication data release

- Providing long-term support of databases

Data producers should state their intentions and enable analyses of their data by:

- Informing data users about the data being generated, data standards and quality, planned analyses, timelines, and relevant contact information, ideally through publication of a citable marker paper near the start of the project or by provision of a citable URL at the project or funding-agency website
- Providing relevant metadata (e.g., questionnaires, phenotypes, environmental conditions, and laboratory methods) that will assist other researchers in reproducing and/or independently analysing the data, while protecting interests

of individuals enrolled in studies focusing on humans

- Ensuring that research participants are informed that their data will be shared with other scientists in the research community
- Publishing their initial global analyses, as stated in the marker paper or citable URL, in a timely fashion
- Creating databases designed to archive all data (including underlying raw data) in an easily retrievable form and facilitate usage of both pre-processed and processed data

Data analysts/users

should freely analyse released prepublication data and act responsibly in publishing analyses of those data by:

- Respecting the scientific etiquette that allows data producers to publish the first global analyses of their data set
- Reading the citeable document

associated with the project

- Accurately and completely citing the source of prepublication data, including the version of the data set (if appropriate)
- Being aware that released prepublication data may be associated with quality issues that will be later rectified by the data producers
- Contacting the data producers to discuss publication plans in the case of overlap between planned analyses
- Ensuring that use of data does not harm research participants and is in conformity with ethical approvals

Scientific journal editors

should engage the research community about issues related to prepublication data release and provide guidance to authors and reviewers on the third-party use of prepublication data in manuscripts

release meetings were not revisited, as they were considered fundamental to all types of data release (whether prepublication or publication-associated). These included: specified quality standards for all data; database designs that meet the needs of both data producers and users alike; archiving of raw data in a retrievable form; housing of both 'finished' and 'unfinished' data in databases; and provision of long-term support for databases by funding agencies. New issues that were addressed include the importance of simultaneously releasing metadata (such as environmental or experimental conditions and phenotypes) that will enable users to fully exploit the data, as well as the complexities associated with clinical data because of concerns about privacy and confidentiality (see 'Sharing data about human subjects', overleaf).

At a practical level, the Toronto meeting developed a set of suggested 'best practices' for funding agencies, for scientists in their different roles (whether as data producers, data analysts/users, and manuscript reviewers), and for journal editors (see 'The Toronto statement').

Recommendations for funders

Funding agencies should require rapid prepublication data release for projects that generate data sets that have broad utility,

are large in scale, are 'reference' in character and typically have community 'buy-in'. The table opposite provides examples of projects using different designs, technologies, and approaches that have several of these attributes, but also lists projects that are more hypothesis-based for which prepublication data release should not be mandated.

It was agreed at the meeting that the requirements for prepublication data release must be made clear when funding opportunities are first announced and that proactive engagement of funders is beneficial throughout a project, as has been the experience of many genome-sequencing

efforts, the International HapMap Project, the ENCODE project, the 1000 Genomes project and, more recently, the International Cancer Genome Consortium, the Human Microbiome Project and the MetaHIT project.

For all projects generating large data sets, the Toronto meeting recommended that funding agencies require that data-sharing plans be presented as part of grant applications and that these plans are subjected to peer review. Such practice is currently the exception rather than the rule. Funding agencies will need to exercise flexibility by, for example, recognizing that large-scale data-generation projects need not necessarily lead to traditional publications, and

that certain projects may need to release only some of their generated data before publication. At the same time, general consistency in data-sharing policies between funding agencies is desirable, whenever possible. To encourage compliance, funding agencies and academic institutions should give credit to investigators who adopt prepublication data-release practices, one option would be to recognize good data-release behaviour during grant renewals and promotion processes, another would be to track the usage and citation of data sets using electronic systems similar to those used for traditional publications⁸.

Data producers and data users

Early data release can lead to tensions between the interests of the data-producing scientists who request the right to publish a first description of a data set and other scientists who wish to publish their own analyses of the same data. To date, many papers have been published by third parties reporting research findings enabled by data sets released before publication. The experiences shared in Toronto suggest that these have rarely affected subsequent publications authored by the data producers. Nevertheless, the Toronto meeting participants recognized that this is an ongoing concern that is best addressed by fostering a scientific culture that encourages transparent and explicit cooperation on the part of data producers, data

"Funding agencies should require rapid prepublication data release for certain projects."

analysts, reviewers and journal editors.

Data producers should, as early as possible, and ideally before large-scale data generation begins, clarify their overall intentions for data analysis by providing a citable statement, typically a 'marker paper', that would be associated with their database entries. This statement should provide clear details about the data set to be produced, the associated metadata, the experimental design, pilot data, data standards, security, quality-control procedures, expected timelines, data-release mechanisms and contact details for lead investigators. If data producers request a protected time period to allow them to be the first to publish the data set, this should be limited to global analyses of the data and ideally expire within one year.

If the citable statement is a 'marker paper' it should be subjected to peer review and published in a scientific journal. Alternatively, other citable sources, such as digital object identifiers to specific pages on well-maintained funding agency or institutional websites, could also be used. Data producers benefit from creating a citable reference, as it can later be used to reflect impact of the data sets⁸.

In turn, the data users should carefully read the source information, including any marker papers, associated with a released data set. Data analysts should pay particular attention to any caveats about data quality, because rapidly released data are often unstable, in that they may not yet have been subjected to full quality control and so may change. It would be prudent for data analysts to assess the benefits and potential problems in immediately analysing released data. They should communicate with data producers to clarify issues of data quality in relation to the intended analyses, whenever possible. In addition, data users should be aware that some data sets are associated with version numbers: the appropriate version number should be tracked and then provided in any published analyses of those data.

Resulting papers describing studies that do not overlap with the intentions stated by the data producers in the marker paper (or other citable source) may be submitted for publication at any time, but must appropriately cite the data source. Papers describing studies that do overlap with the data producer's proposed analyses should be handled carefully and respectfully, ideally including a dialogue with the data producer to see if a mutually agreeable publication schedule (such as co-publication or inclusion within a set of companion papers) can be developed. In this regard, it is important

Sharing data about human subjects

Data about human subjects participating in genetic and epidemiological research require particularly careful consideration owing to privacy-protection issues and the potential harms that could arise from misuse. These issues are critical to all databases housing information about human subjects, whether or not they contain prepublication data. For these reasons, it is

important to develop and implement robust governance models and procedures for human subjects data early in a project. Lessons can probably be learned from data policies adopted by several genomics projects⁹ that generate human-subject data. For aggregated data that cannot be used to identify individuals, databases are open access, but for clinical and genomic data that are

associated with a unique, but not directly identifiable individual, access may be restricted.

Under such conditions, arguments can be made for the release of data for studies involving human subjects, as doing so can augment the opportunities for new discoveries that could ultimately benefit individuals, communities, and society at large.

for data users to realize that, historically, many such dialogues have led to coordinated publications and to new scientific insights. Despite the best intentions of all parties, on occasion a researcher may publish the results of analyses that overlap with the planned studies of the data producer. Although such instances are hopefully rare if good communication protocols are followed, these should be viewed as a small risk to the data producers, one that comes with the much greater overall benefit of early data release.

Editors and reviewers

As reviewers of manuscripts submitted for publication, scientists should be mindful that prepublication data sets are likely to have been released before extensive quality control is performed, and any unnoticed errors may cause problems in the analyses performed by third parties. Where the use of prepublication data is limited or not crucial to a study's conclusions, the reviewers should only expect the normal scientific practice of clear citation and interpretation. However, when the main conclusions of a study rely on a prepublication data set, reviewers should be satisfied that the quality of the data is described and taken into account in the analysis.

Participants at the Toronto meeting recommended that journals play an active part in the dialogue about rapid prepublication data release (both in their formal guide to authors and informal instructions to reviewers). Journal editors should remind reviewers that large-scale data sets may be subject to specific policies regarding how to cite and use the data. Ultimately, journal editors must rely on their reviewers' recommendations for reaching decisions about publication. However, encouraging reviewers to carefully check the conditions for using data that authors have not

created themselves can help to raise both the quality of analysis and fairness in citation of published studies.

Conclusion

The rapid prepublication release of sequencing data has served the field of genomics well. The Toronto meeting participants acknowledged that policies for prepublication release of data need to evolve with the changing research landscape, that there is a range of opinion in the scientific community, and that actual community behaviour (as opposed to intentions) need to be reviewed on a regular basis. To this end, we encourage readers to join the debate over data-sharing principles and practice in an online forum hosted at <http://tinyurl.com/lqxp3>. ■

Toronto International Data Release Workshop Authors. A complete list of the authors and their affiliations accompanies this article online. e-mail: birney@ebi.ac.uk, tom.hudson@oicr.on.ca

"Prepublication data are likely to be released before extensive quality control is performed."

1. Committee on Responsibilities of Authorship in the Biological Sciences, National Research Council Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences (National Academy of Sciences, 2003).
2. International Human Genome Sequencing Consortium *Nature* **409**, 860–921 (2001).
3. *Summary of Principles Agreed at the First International Strategy Meeting on Human Genome Sequencing* Bermuda, 25–28 February 1996 (HUGO, 1996); available at www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml
4. *Summary of the Report of the Second International Strategy Meeting on Human Genome Sequencing* Bermuda, 27 February–2 March 1997 (HUGO, 1997); available at www.ornl.org/sci/techresources/Human_Genome/research/bermuda.shtml#2
5. Guyer, M. *Genome Res.* **8**, 413 (1998).
6. *Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility* (Wellcome Trust, 2003); available at www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtd003207.pdf
7. Rodriguez, H. et al. *J. Proteome Res.* **8**, 3689–3692 (2009).
8. *Nature Biotechnol.* **27**, 579 (2009).
9. Kaye, J., Heeney, C., Hawkins, N., de Vries, J. & Boddington, P. *Nature Rev. Genet.* **10**, 331–335 (2009).

Join the discussion at <http://tinyurl.com/lqxp3>

See online special at <http://tinyurl.com/dataspecial>

Post-publication sharing of data and tools

Despite existing guidelines on access to data and bioresources, good practice is not widespread. A meeting of mouse researchers in Rome proposes ways to promote a culture of sharing.

Sharing scientific data through publication has long underpinned the cycle of discovery and is the dominant means by which scientists earn credit for their work. More recently, technologies generating very large data sets and novel biological materials have given rise to principles under which communities share data and materials (pre- and post-publication), and to a new sharing infrastructure — large public databases and repositories. Although much attention has been given to practical and ethical guidelines for prepublication data release from large-scale ‘community resource projects’, summarized in the Bermuda Principles¹ and the Fort Lauderdale report², sharing of data and resources from hypothesis-driven research has largely been addressed piecemeal by individual communities, journals and funding agencies.

We report here the efforts of one such community to address issues of particular relevance to the free sharing of data and resources for mouse biology, genetics and functional genomics. Our community has had more than six decades experience with strategies for sharing mice, and more recently for cell lines. When it comes to resource sharing, the two greatest impediments to fully exploiting global research using the mouse as a model organism are the barriers created by material transfer agreements and the underutilization of public mouse repositories.

Community discussion

At a meeting in Rome in May organized by the CASIMIR consortium, a European project examining mouse research infrastructure, participants attempted to establish an agenda for community discussion. This meeting was attended not just by mouse investigators, but by representatives of funding agencies and journals, intellectual-property specialists and sociologists. The resulting Rome Agenda was designed to assist the stakeholders in developing a coordinated and directed approach to the main factors inhibiting free sharing of the fruits of publicly funded mouse research.

Two of the most important shared resources and research outputs in the field are mice and embryonic stem cells. The imperative to share such resources was probably first articulated by the US National Institutes of Health (NIH) in

March 1984. Yet even today, numerous unique mouse strains are not made available to the research community despite the existence of publicly funded mouse repositories provided for this purpose (see International Mouse Strain Resource (IMSR), www.findmice.org). Comparison of the number of knockout mice recorded by the international Mouse Genome Informatics (MGI) database (<http://www.informatics.jax.org/>) with those deposited in IMSR repositories suggests that currently only 35% are available in this way. This is an encouraging doubling of the percentage available since last assessed in a 2006 NIH survey. To further improve this figure, however, it is important that the sharing ethos is consistently observed by the mouse community and investment in repositories continues to keep pace with the generation of new strains.

Experiences shared at the meeting indicated that enforcement of existing policies regarding data and resource deposition is variable, and that despite increased emphasis on the importance of sharing by journals and funding organizations in recent years, there is evidence that geneticists and genomic researchers are withholding data and research materials with increasing frequency³. It is one thing

"Enforcement of existing policies regarding data and resource deposition is variable."

to encourage data deposition and resource sharing through guidelines and policy statements, and quite another to ensure that it happens in practice, as a recent informal survey of proteomics data deposition has revealed⁴.

Consequently, although many of the issues discussed in Rome are of specific concern to mouse biology and functional genomics, several have relevance to the wider biological sciences. For example, the issues surrounding licensing and patenting of genetically manipulated mice and embryonic stem cells could apply to many research tools that are generated through hypothesis-driven research. We hope that our experiences and recommendations can inform and stimulate broad discussion in the community as a whole and we ask readers to participate in an online forum to that end (see <http://tinyurl.com/mo4gh8>).

A strong message from Rome was that funding organizations, journals and researchers need to develop coordinated policies and actions on sharing issues. The Rome Agenda described

and summarized here (see ‘The Rome Agenda, overleaf’), represents a challenge to stakeholders to coordinate their efforts to facilitate the ready exchange of data and resources and to share good practices already implemented by some organizations and journals.

Access to publication-associated data

Prepublication data release is comprehensively discussed in an accompanying paper from the Toronto group⁵, whose conclusions were broadly supported in Rome. For publication-associated data, the meeting strongly endorsed the recommendations of the National Academy of Sciences UPSIDE report⁶, which lays out detailed guidelines for data sharing, not least the principle that data on which publications are based should be made available immediately on publication.

Currently, funding bodies rarely require investigators to deposit their mice in public repositories, although many encourage it, with the consequence that mutant lines may be lost or not fully exploited. The meeting strongly recommended that, at least on publication, journals should insist that mice and embryonic stem cells be deposited in a public repository within a specified time frame, the length of which still requires community consensus. Additionally, funders should be willing explicitly to cover the costs of deposition of mice arising from projects into public repositories.

We recommend that it becomes mandatory for scientific papers to explain where and how to access data and resources generated as part of the investigation. We are aware that some journals already have strong policy positions in this area, insisting that large data sets must be deposited in public databases, and that all reasonable requests for materials from other researchers must be fulfilled. There is however, heterogeneity with both policy and enforcement; surprisingly, many journals have no written policy on the availability of either bioresources or primary data.

In addition, papers should acknowledge any other data or materials used and the originating sources. This might be facilitated by the addition of metadata tags linking to data and bioresources⁴. A mechanism, such as a digital object identifier for resources in public repositories, would allow ready searching of the literature for specific bioresources, which is currently extremely difficult. It

would also add incentives for complying with data release and deposition policies by attributing credit to researchers who do share.

When it comes to compliance, journals and funding agencies have the most important role in enforcement and should clearly state their distribution and data-deposition policies, the consequences of non-compliance, and consistently enforce their policy. The costs of proactive 'policing' (explicit review at the end of grants or following publication) may be disproportionate, but a consistently implemented reactive policy, in a culture in which sharing is the ethical norm would, we believe, suffice.

Where they don't yet exist, clear criteria should be developed for reviewers of grants to help them assess data and material-sharing plans submitted as part of a funding proposal. There are already examples of good practice in this regard from the NIH⁷, the Howard Hughes Medical Institute, and several UK funding organizations such as the Wellcome Trust and the Medical Research Council⁸⁻¹⁰. Data-sharing plans are required in proposals, efforts are made to facilitate sharing, such as putting investigators in touch with repositories and, for some organizations, compliance is an important consideration in funding renewal.

Deposition of data and resources into public repositories is important for the validation of published results, as well as facilitating reuse. Although it is usual practice for major public databases to make data freely available to access and use, any restrictions on use should be strongly resisted and we endorse explicit

encouragement of open sharing, for example under the newly available CC0 public domain waiver of Creative Commons¹¹.

Licensing, patenting and material transfer agreements

Recent experience from technology-transfer programmes in the public sector discussed at the Rome meeting reflects a growing consensus among technology-transfer professionals that the patenting of mouse resources and genes is expensive and a poor return on investment. (Not least because most research tools are available under non-exclusive licences, whether patented or not.) This is reflected in a 1999 NIH policy that discourages filing of patents on mice as research tools generated from work done in its intramural research programmes. We recommend patenting research tools and methods only under exceptional circumstances, although patents may still be appropriate for research methods that are broadly applicable to multiple research fields.

Regardless of whether mouse resources or research methods are patented, licensing terms should be as broad as possible, acknowledging that academic institutions are both developer-providers and recipient-users of new mouse models, so there is little benefit in imposing obstacles on the availability and use of mice in the form of patents, licences and material transfer agreements (MTAs). Moreover, researchers

should be free to breed these mice for internal research purposes and to cross-breed them to develop innovative new mouse models.

With commercial use, any licensing of mice or methods to the private sector should include a broad reservation of rights on behalf of academic and not-for-profit institutions to use the mouse or method for non-commercial research purposes. In accordance with the sharing policies of some funding institutions, such as the NIH, it would be inappropriate to include licensing terms requiring royalty

reach-through or product reach-through on subsequent inventions, and institutional policies on intellectual property, technology transfer and licensing should reflect these principles. Equally, repositories should be able to distribute mouse resources to industry

under reasonable terms and conditions.

Within the academic community, processing of MTAs has become a major impediment to the open and timely dissemination of mouse resources and associated data¹². Onerous terms and conditions in many MTAs have increased transactional costs for institutions and have become a major cause of delay in negotiations and the sharing of resources. We recommend that materials and data be shared under the least restrictive terms possible. If documentation is necessary for any reason, then the minimum NIH sharing policy should be applied¹³. This 1999 policy states that materials developed

"We recommend that materials and data be shared under the least restrictive terms possible."

The Rome Agenda

Access to data and materials

- The data on which publications are based should be made available immediately through public databases on publication. Journals should insist that mice or embryonic stem cells are deposited in a public repository within a specified time frame.
- It should become mandatory for publications to explain where and how to access data and materials generated during the investigation. Publications should acknowledge any other data or materials used, the originating sources and availability.
- Grant reviewers should be provided with clear guidelines to assess data- and material-sharing plans, whether these have been met in the application, and whether the mechanism of sharing proposed would meet appropriate

goals if the work was to be funded or ultimately published.

- Funding organizations should be willing explicitly to cover the costs of deposition of materials arising from projects as part of the project budget.

Licensing and patenting

- The public sector should patent mice as research tools only under exceptional circumstances.
- Licensing terms for mouse resources or research methods should promote the establishment of a mouse 'research commons'.
- Materials and data should be shared under the least restrictive terms possible. Material transfer agreements for sharing materials between academic and not-for-profit institutions should be avoided or simplified.
- Researchers should be free to

breed shared mice for internal research purposes and to cross-breed to develop new mouse models.

- Licensing of mice or methods for commercial use should include a broad reservation of rights for academic and not-for-profit institutions.
- Licensing terms should not include inappropriate royalty reach-through or product reach-through on subsequent inventions, and institutional policy should reflect this.

Data and resource-sharing infrastructure

- Further dedicated sustainable investment in public databases and repositories should be encouraged.
- Funding agencies should provide researchers with clear direction on

expectations for data/resource/publication sharing, and should ensure appropriate data-sharing plans at the outset of projects and facilitate sharing as data and resources are generated.

Standards and tool development

- Data structure and semantics need standardizing and adopting.
- Metadata should be consistently attached.
- Investment is needed in computational tools to make use of standards and interoperability for data sharing and reuse.

Attribution and reward

- Attribution of data or resources should be enforced by journals and databases.
- A system for measuring attribution is needed to provide rewards for data sharing.

using NIH Federal funding should be freely transferred between researchers using "... either no formal agreement, a cover letter, the Simple Letter Agreement of the Uniform Biological Materials Transfer Agreement (UBMTA), or the UBMTA itself".

The Jackson Laboratory in Bar Harbor, Maine, an example of good practice, has applied these principles for many years. The laboratory provides mice to academic and not-for-profit researchers with the simple notification that the mice are to be used solely for research purposes and are not to be sold or transferred to third parties without permission.

Data and resource-sharing infrastructure

The view of meeting participants was that the largest part of the data underlying publications is archived on journals' 'supplemental information' sites or authors' own sites. These data are often formatted in a non-standard way, not readily searchable, and in the long term not guaranteed to persist. In a 2006 survey of major journals, Anderson *et al.*¹⁴ found that on average only 83% of supplementary data were still accessible a year after publication (for one journal this was as low as 33%) and that it seemed that approximately 10% of all data that was supposed to be available through a supplementary website was never available at all. It is clear, therefore, that the issue of long-term sustainable public repositories needs to be addressed by funding agencies, publishers and the community.

Many of the major public data repositories have no stable underlying funding and there are data types, particularly new ones, without appropriate public data repositories. We encourage further investment and recommend that public database coverage and stability be looked at in a coordinated way by funding organizations and the community with increased urgency. A good model is provided by the UK Biotechnology and Biological Sciences Research Council's Bioinformatics and Biological Resources Fund, which provides dedicated funding for development and sustainability of public resources and informatics tools.

Standards and tool development

Shared data are useful only if they are searchable and usable. For both attributes data must be formatted in a standard way, conform to standard structure and semantics and have appropriate metadata attached. It is clear that the community is still a long way from achieving these standards; further support and community discussion is needed. The full utility of standards such as MIBBI (Minimum Information for Biological and

Research commons

A research commons is a set of resources available to all scientists, either as part of the public domain or on standard terms and conditions that facilitate scientific collaboration, efficient reuse of materials and data, and dissemination of knowledge.

Biomedical Investigations) will be attained only by developing tools for data retrieval, mining and computation. The Gene Ontology bioinformatics initiatives provide a good example of how parallel development of tools and standards generates added value. Dedicated funding is needed to develop key elements of database infrastructure, including interoperability and data integration.

Common agenda

Despite oft-repeated statements of good intentions, stakeholders do not always share common interests. Within academia, a fear of 'helping the opposition' runs alongside concerns about the ethical or responsible use of freely shared data. A culture of sharing and open access is made more difficult by policies promoting the commercialization of research¹⁵, ineffective sharing infrastructure and inadequate data standards. Combined with unrealistic expectations from institutions of the value of exclusive licensing to the highest bidder, these factors can slow the progress of discovery and translation.

As an antidote to these concerns, the Rome meeting strongly encouraged sharing behaviours that promote a 'research commons' (see box, above). The heart of a research commons is one in which academic research is not impeded by restrictions on use and access to data and materials, in line with the principles of the Creative Commons¹¹. Adoption of a set of 'mouse research commons' principles would increase the effective use and economic value of publicly funded research by avoiding duplication of effort, unnecessary creation and use of live animal models, and facilitating reuse of data.

We know from the Jackson Laboratory's experience with its repository that developers of new mouse resources are willing to comply with an unrestrictive distribution policy as a condition for acceptance of their resources, so we believe the mouse research commons is not just a utopian dream. Rather it should create a paradigm shift to establish this as a norm for the research community. ■

Paul N. Schofield is in the Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, CB2 3EG, UK, **Tania Bubela** is in the Department of Public Health Sciences, University of Alberta, Edmonton Alberta, T6G 2V2, Canada, **Thomas Weaver** is

at the MRC Mary Lyon Centre, Harwell, Didcot, Oxfordshire, OX11 0RD, UK, **Lili Portilla** is at the National Center for Research Resources, Bethesda, Maryland 20892-4874, USA, **Stephen D. Brown** and **John M. Hancock** are at MRC Harwell, Mammalian Genetics Unit, Harwell Science and Innovation Campus, OX11 0RD, UK, **David Einhorn** is at the Jackson Laboratory, Bar Harbor, Maine 04609, USA, **Glaucio Tocchini-Valentini** is at the Istituto di Biologia Cellulare, 00015 Monterotondo Scalo, Rome, Italy, **Martin Hrabe de Angelis** is at the Institute of Experimental Genetics, Munich, Germany, **Nadia Rosenthal** is at the EMBL Monterotondo, 00015 Monterotondo, Rome, Italy, for the **CASIMIR Rome Meeting participants**. A complete list of the authors and their affiliations accompanies this article online.

e-mail: PS@mole.bio.cam.ac.uk

1. Summary of Principles Agreed at the First International Strategy Meeting on Human Genome Sequencing Bermuda, 25-28 February 1996 (HUGO, 1996); available at www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml
2. Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility (Wellcome Trust, 2003); available at www.wellcome.ac.uk/Stellent/groups/corporatesite/@policy_communications/documents/web_document/wtd003207.pdf
3. Cohen, W. M. & Walsh, J. P. *Innov. Policy Econ.* **8**, 1-30 (2008).
4. *Nature Biotechnol.* **27**, 579 (2009).
5. Toronto International Data Release Workshop Authors *Nature* **461**, 168-169 (2009).
6. Committee on Responsibilities of Authorship in the Biological Sciences, National Research Council Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences (National Academy of Sciences, 2003).
7. Final NIH Statement on Sharing Research Data, NOT-OD-03-032 (National Institutes of Health, 2003); available at <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>
8. Medical Research Council Policy on Data Sharing and Preservation; available at www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Datasharinginitiative/Policy/index.htm
9. Wellcome Trust Policy on Data Management and Sharing; available at www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm
10. Biotechnology and Biological Sciences Research Council Data Sharing Policy; available at www.bbsrc.ac.uk/publications/policy/data_sharing_policy.html
11. Science Commons Database Protocol <http://sciencecommons.org/resources/faq/database-protocol>
12. Walsh, J. P., Cohen, W. M. & Cho, C. *Res. Policy* **36**, 1184-1203. (2007).
13. Principles and Guidelines for Recipients of NIH Research Grants and Contracts on Obtaining and Disseminating Biomedical Research Resources Federal Register **64**, 72090-72096 (1999); available at http://grants.nih.gov/grants/intell-property_64FR72090.pdf
14. Anderson, N., Tarczy-Hornoch, P. & Bumgarner, R. E. *BMC Bioinformatics* **7**, 260 (2006).
15. Nelson, R. R. *Res. Policy* **33**, 455-471 (2004).

The views expressed in this paper represent the consensus of the meeting participants, and do not necessarily reflect the current policy of their respective Institutions.

Join the discussion at <http://tinyurl.com/mo4gh8>

See online special at <http://tinyurl.com/dataspecial>

BOOKS & ARTS

Call for a climate culture shift

A new book describes the rapid reshaping of human priorities needed to save the planet from global warming. Some of that change is already under way at the community level, explains **Robert Costanza**.

**Down to the Wire:
Confronting Climate Collapse**

by David W. Orr

Oxford University Press: 2009. 288 pp.
\$19.95

In the fight against climate change, humans will need to do more than switch to energy-efficient light bulbs and buy 'green' goods. As environmental scientist David Orr points out in *Down to the Wire*, what is needed is a radical shift in culture that alters our priorities. The question is whether that task, which seems impossible, can be made to happen. Orr's book, along with recent research and social initiatives, give hope that it can.

There is a growing scientific consensus that humanity is rapidly approaching a global climate catastrophe. Although we have increasing knowledge of the dangers and costs ahead there is little time to avert a disaster. Orr acknowledges these dire circumstances, but does not wallow in despair or defeatism. His book is a clear-sighted view of what we need to change now.

He describes three essential categories of radical change, in increasing order of difficulty. The first and most easily achievable is a redesign of the infrastructure for producing food, energy, water and other commodities so that it is powered by renewable sources. Second is an overhaul of education systems to develop ecological literacy and encourage creative, real-world problem solving. The third is to reform our political systems from the current corporate plutocracies to true democracies with real leaders.

We live, as Orr says, "amid the ruins of failed '-isms'". Both communism and capitalism have pursued policies of 'growth at all costs' that have failed to account adequately for the value of natural and social capital assets, such as a stable climate, functioning ecosystems

and successful human communities. An alternative solution is needed to formulate a fundamentally different set of economic goals for society. Orr prescribes three such goals "that presently appear to be utterly impossible". First, he advocates a change in priority: instead of economic growth, we should switch to development that genuinely improves the

"A worldwide movement rejects the idea that we are fated to end the human experiment with a bang or a whimper."



Many communities choose to build sustainably, such as the solar-powered Lewis Center in Oberlin, Ohio.

quality of life for everyone. Second, consumer culture should be focused on needs, not wants; and third — hardest of all — we should summon "the compassion and wisdom to fairly distribute wealth, opportunity, and risk".

These goals and the policies to achieve them have long been on social and political agendas. Why have things not changed, and how can they be made to change? Orr addresses this adroitly, showing that human nature is flexible and that rapid cultural shifts have happened before. In the United States after the Second World War, for example, the culture changed to allow new social and taxation policies that created the middle class. The rapid fall of the Soviet Union resulted from the slow build-up of social problems until a tipping point was reached. It may only be a matter of time before people who share the goals of quality of life, fairness and sufficiency begin to outweigh those whose world view is locked into growth at all costs.

Some evidence that such attitudes are on the rise comes from the work of sociologist Paul H. Ray and psychologist Sherry Ruth Anderson, who have surveyed and categorized world views in the United States over the past four decades. In their book *The Cultural Creatives*

(Crown Publishing; 2000), they break the US population into three groups: 'traditionals', who include the religious right and others who hark back to the past; 'moderns', who are the current dominant group and include the 'growth at all costs' type; and 'cultural creatives', including those with the values and goals that Orr promotes. The percentage of cultural creatives in the United States increased from almost nothing in the 1960s to 25% by the year 2000, and is now close to 30% by some estimates. A political tipping point will occur when this percentage is large enough to begin to radically change the political dynamics of the country and, by extension, the world.

As Orr points out, many varied initiatives are already pressing towards a cultural shift. Examples include the 'transition town' movement, spearheaded by the charity Transition Network in Totnes, UK, which aims to help communities reduce their carbon emissions; the 'sustainable cities' effort based in Vancouver, Canada, which supports urban sustainability projects worldwide; and Orr's own initiative to plan and construct sustainable buildings in the city of Oberlin, Ohio.

Other indicators of this shift include the thousands of organizations that are devoted to restoring the environment and fostering social justice, as described by Paul Hawken in his 2007 book *Blessed Unrest* (Viking).

B. TAXEL/WILLIAM McDONOUGH + PARTNERS

And a French government commission, set up in 2008 to assess economic performance, is one of many attempts to account for the limitations of the gross domestic product as a measure of social progress. Such examples are evidence of the growing global dialogue on providing real solutions to the problem of building a sustainable and desirable future. A journal entitled *Solutions* (of which I am editor-in-chief, and Orr and Hawken are associate editors) is due to launch soon to add to these discussions.

All of this shows that a global cultural shift

and transformation is indeed in progress. As Orr concludes, this transformation “has grown into a worldwide movement that rejects the idea that we are fated to end the human experiment with a bang or a whimper on a scorched and barren Earth.” We still have a choice, but it is now or never. Orr’s book will do much to help achieve the required cultural transformation, hopefully just in time. ■

Robert Costanza is director of the Gund Institute for Ecological Economics at the University of Vermont, Burlington, Vermont 05405, USA. e-mail: robert.costanza@uvm.edu

mathematics. Tett, who runs the coverage of global markets for the *Financial Times* newspaper, came to journalism after taking a doctorate in social anthropology. Her book describes how a small group of young bankers with training in mathematics, physics and actuarial science created a class of financial products known as credit derivatives, the misuse of which helped to precipitate today’s crisis.

Tett recounts how a group at J. P. Morgan, one of the United States’ oldest commercial banks, came up with a scheme designed to free up more of the bank’s capital for profit-making investment: selling the loans on their books to third-party buyers. The regulators were unsure whether this was allowed, but Tett shows how they were eventually won over following an aggressive lobbying campaign led by the bank.

Interestingly, J. P. Morgan decided not to shift bundles of mortgage loans in the same way as they were selling off commercial loans. This is because the bank did not have enough data to accurately predict the numbers of borrowers who would default, or the extent to which one defaulter might trigger others. Other banks, however, began to offer bundles of individual mortgages — including those given to ‘sub-prime’ clients. This practice was helped along by David Li, an actuarial scientist who published a formula that claimed to predict patterns of defaulting without needing data on individual financial histories. The market for sub-prime loan bundles went through the roof, and when these sub-prime borrowers began to default, the world economy went through the floor.

Tett concludes that banking needs to go back to an earlier philosophy: products should be simple to understand, bankers need to respect the fact that regulators are acting in the public interest, and regulators need to be more on the ball. This checklist sounds like the regulatory hurdles that many scientists have experience of — especially

those working in areas such as food safety and environmental protection.

For example, after the bovine spongiform encephalopathy (‘mad cow’) epidemic and controversies over genetically modified crops in the 1990s, there was much soul-searching in Britain over how best to protect the public from possible future harm. One outcome was

The wider lessons for finance

One of the unintended effects of the near-collapse of the world economy is the creation of a market for scientific advice to the banking sector. Senior officials at the Bank of England, for example, are consulting the theoretical ecologist and former Royal Society president Robert May, whose research interests include modelling ecosystem collapses and the spread of infectious diseases. Why? Because May’s work could provide signposts on how to develop a comprehensive model for the movement of money around the world and the myriad connections between cash, individuals and institutions.

Two books highlight insights from other fields, such as psychology and anthropology, on the current global financial situation, and the damage done in the past by inappropriately applied mathematics. They also suggest that finance can learn even greater lessons from science, by taking account of the experiences of scientists and science’s ideal of transparency and regulation.

In *Animal Spirits*, two Keynesian economists — George Akerlof, a Nobel-prizewinning economist at the University of California, Berkeley, and Robert Shiller, an economist at Yale University — use findings from psychology to amplify one of economist John Maynard Keynes’s theories. In his signature 1936 work, *The General Theory of Employment, Interest and Money*, Keynes explained that economies should fluctuate because people behave in unpredictable ways — under the influence of what he called “animal spirits”.

Keynes’s theory countered the mainstream view in economics that people, and therefore markets, behave in rational ways. But people often decide which house to buy or which car to drive because it ‘feels right’, or for other

reasons that economists may find irrational or cannot measure accurately. Akerlof and Shiller remind us that emotional and intangible factors — such as confidence in institutions, illusions about the nature of money or a sense of being treated unfairly — can affect

how people make decisions about borrowing, spending, saving and investing.



Animal Spirits is an affectionate tribute to the man whose ideas, unfashionable for the past 30 years, have resurged. Having advised governments through the Depression, Keynes became convinced that

more government spending was needed to maintain employment during a recession — a prescription that has been adopted by many national leaders, including UK Prime Minister Gordon Brown and US President Barack Obama.

What *Animal Spirits* doesn’t do is illustrate how descriptions of human behaviour can be used in quantitative financial and economic models. Bankers, analysts and policy-makers reading the book will want to know where they can find data on human behaviour, and how

these data can be reduced to the indices, constants and variables that make up their equations. Akerlof and Shiller would have done well to have included a chapter covering this issue.

Insights from science have the potential to be misused, however. Gillian Tett’s book *Fool’s Gold* is an exceptional account of how today’s financial world became in thrall to advanced

Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism

by George A. Akerlof and Robert J. Shiller
Princeton University Press: 2009.
264 pp. £16.95

Fool’s Gold: How Unrestrained Greed Corrupted a Dream, Shattered Global Markets and Unleashed a Catastrophe

by Gillian Tett
Little, Brown: 2009. 352 pp. £18.99

a new food-standards agency, independent of government and industry, to act in the interests of consumers. Many scientists were not enthusiastic about this change. But those who have worked in the system realize that good regulation and transparency are not enemies of progress.

The banking industry is waking up to

the fact that advanced knowledge helped to create profits beyond imagination, but that greed and secrecy played a part in its near-downfall. *Animal Spirits* gives hope that such knowledge can be a force for good. *Fool's Gold*, meanwhile, reminds us that this must go hand-in-hand with transparency and keeping the public interest uppermost. ■

Ehsan Masood teaches international science policy at Imperial College London.
e-mail: em@ehsanmasood.com

Ehsan Masood will chair a *Nature* debate on science and the financial crisis in London on 21 September — details at <http://tinyurl.com/nc9pvn>. For more on the economy, see <http://tinyurl.com/nlb76n>.

How Spain redrew the world

Secret Science: Spanish Cosmography and the New World

by María M. Portuondo

University of Chicago Press: 2009.

360 pp. \$45

In the autumn of 1571, Juan López de Velasco, an ambitious legal scholar with one eye on the heavens, accepted the coveted position of chief cosmographer and chronicler to Philip II, the King of Spain. Velasco received a salary hike and a trunk filled with invaluable documents collected by his predecessor. In the years that followed, the maps, treatises and narrative accounts found inside the trunk revealed the geography of a new world to this enthusiastic map-maker, whose job included updating the empire's navigational charts and keeping ships' pilots and government bureaucrats informed of any new geographical data retrieved from overseas. Velasco sat at the centre of one of the most successful information-gathering operations the world had ever known. But his work remained secret for centuries.

Thanks to several recent studies, the private knowledge held by generations of servants to the Iberian Crown — and hidden for centuries in musty archives — has now been thrust into the public eye. One such study is María Portuondo's impeccably researched book on Spanish cosmographical practice. Cosmography was a discipline that involved creating textual descriptions of the known world using charts and images similar to those bequeathed to Velasco. The early modern equivalent of satellite-enhanced telemetry, these colourful cartographic images served sixteenth-century monarchs and their ministers in pragmatic ways. They were used for plotting trade routes, tracing the design of new cities, conceiving military campaigns and imagining the world's emerging political boundaries.

At stake in these materials was the very nature of scientific practice itself. Long before Galileo Galilei cast doubt on the existence of an Earth-centred Universe, Spanish navigators and royal cosmographers were already working to overturn centuries of received wisdom about the layout of the cosmos and Earth's place within it. Science, and the direction of the modern world, would never be the same.

Cosmography was Renaissance shorthand for several modern disciplines: astronomy, history, geography, anthropology, navigation and the study of nature. Practitioners brought together these techniques with the classical

New World and a crescendo of reports from Spanish pilots heralding islands and continents that were previously unknown, cosmographers had to quickly adjust the master chart of the world, held at the House of Trade in Seville, Spain's central clearing house for geographical information. Eyewitness observations from the Americas would trump ancient theories as the world map was redrawn.

In the past two decades or so, some Spanish historians of science have adopted a defensive tone when discussing the supposed lack of Iberian prowess during the scientific revolution — as compared with the better-known discoveries made by northern Europeans such as Francis Bacon, Johannes Kepler or Isaac Newton. They argue that Spanish and sometimes Portuguese navigators were precursors to those 'revolutionary' scientific activities. Spanish pursuits in astronomy, navigation and other empirical disciplines, they assert, have historically been neglected or ignored owing to long-standing prejudice and misinformation.

Particularly refreshing in Portuondo's tale is the absence of such an attitude. Rather, she shows how a cast of eclectic men of letters in service to the Spanish crown set out to change the image of the world. They developed elaborate geographical questionnaires to learn from local populations, and sponsored programmes of celestial observations — during lunar eclipses, for instance — to make a global network of field laboratories out of their monarch's territorial possessions, even though the results were never publicized beyond a privileged few.

Similarly to Velasco's chest of cartographic treasures, Portuondo's study reveals valuable evidence with which scholars can refashion their images of the Renaissance world and the achievements of Spanish science at the dawn of modernity. ■

Neil Safier teaches history at the University of British Columbia, Vancouver V6T 1Z1, Canada, and is author of *Measuring the New World*.
e-mail: neil.safier@ubc.ca



Discovery of the New World stretched geographical boundaries and Spanish cosmographers' skills, exemplified in this 1580 map of Oaxtepec, Mexico.

and biblical narratives describing the shape of the known world, which, at the time, stretched barely beyond the African coast and Asia minor — as described by influential geographers such as the Greek Strabo and the Roman Pomponius Mela.

But in 1492 everything changed. Once Spanish galleons had crossed the Atlantic under the command of the Genoese navigator Christopher Columbus, Spanish cosmographers were forced to reconcile increasingly contradictory ways of thinking. With the discovery of the

NEWS & VIEWS

SEX DETERMINATION

Birds do it with a Z gene

Jennifer A. Marshall Graves

The gene that determines sex in birds has eluded scientists for a decade. Now this all-important locus is revealed as a gene on the Z chromosome known for its proclivity for determining sex in all kinds of animals.

Sex in birds, as well as in humans and other mammals, is determined by genes on specialized sex chromosomes. Mammalian females have two X chromosomes and males a single X and a degenerate Y chromosome that bears the male-dominant testis-determining gene *SRY*. Birds are just the other way around, with males having two Z chromosomes and females a single Z and a W chromosome. The sex-chromosome pairs of most mammals and birds share none of their suites of genes, but the gene-rich bird Z chromosome and the condensed and gene-poor W chromosome show uncanny parallels with the mammalian X and Y chromosomes. The ZW and XY sex-chromosome pairs evidently evolved from different non-sex chromosome (autosome) pairs as one of the two partners acquired a sex-determining locus and then degenerated. There is no sign of a bird *SRY* gene, nor are there convincing candidate female-determining genes on the W chromosome, leaving bird sex determination somewhat up in the air. In this issue (page 267), in work that tests the credentials of a gene on the chicken Z chromosome, Smith *et al.*¹ provide direct evidence that the *DMRT1* gene is the long sought bird sex-determining gene.

Over the years *DMRT1* has been a seductive candidate for bird sex determination. It lies on the Z chromosome and has no copy on the W, even in the emu, in which the W chromosome is similar to the Z. *DMRT1* is transcribed specifically in the testis and is not 'dosage compensated' — it is active on both Z chromosomes in males although only one copy is present in females. So *DMRT1* could function by a dose-related mechanism, in which a threshold amount of gene product is needed to make a testis. This threshold can be reached only by ZZ birds, whereas half this amount is insufficient, leaving the ZW gonad to pursue a default female pathway.

Smith and colleagues' work¹ comes from the laboratory of Andrew Sinclair, a co-discoverer of both human *SRY*² and chicken *DMRT1* (ref. 3). This group had tried for years to knock out, or knock down, or insert extra copies of *DMRT1* into chicken embryos, a messy

task in birds because of their large and yolky eggs. They have now managed to knock down *DMRT1* in chicken embryos using interfering RNA delivered by an avian retroviral vector, with viral spread monitored by expression of a green fluorescent protein (GFP) marker. Knockdown of genes in birds is technically challenging, but out of 550 embryos injected with the virus, 24% showed GFP fluorescence and a lowered expression of *DMRT1* (ref. 1).

The thrilling result was that gonads in ZZ embryos in which *DMRT1* had been successfully knocked down looked more like ovaries than testes. Feminization was convincingly demonstrated by examining the microscopic appearance of gonadal tissue (which showed

disruption of testis cords, among other changes), and by measuring the expression of male and female marker genes (for *SOX9* and aromatase, respectively). The decreased concentration of *DMRT1* protein in the knockdown embryos was evidently unable to support male gonad development. These data strongly support the hypothesis that *DMRT1* is the chicken sex-determining gene and provide direct evidence that it functions by dosage differences between males and females.

How significant is this discovery to our understanding of vertebrate sex? *DMRT1* is involved with sex in other species, as shown by impaired testis development in *Dmrt1*-mutant mice⁴ and by male-to-female sex reversal in XY humans who have a deletion of the end of chromosome 9 (the human counterpart of the chicken Z chromosome⁵). A recently evolved sex-determination system in medaka fish (*Oryzias latipes*) uses a copy of *DMRT1* (known as *DMY*) to define a new Y chromosome⁶. This increases the amount of *DMRT1* above the male-determining threshold⁶, again consistent with the hypothesis that *DMRT1* dosage is crucial for sex determination in all vertebrates.

Even in the fruitfly *Drosophila melanogaster* and the worm *Caenorhabditis elegans*, an equivalent of *DMRT1* is involved in sex; in both species it mediates male mating behaviour as well as regulating transcription of yolk-protein genes and differentiation of male-specific sense organs⁴. In the African clawed frog *Xenopus laevis* a W-borne *DMRT1* copy, *DMW*, is involved in ovarian development⁷, perhaps by inhibiting *DMRT1* action.

How does *DMRT1* do its job? It encodes a protein containing a cysteine-rich DM domain that, at least in invertebrates, transcriptionally regulates target genes. In birds, it probably heads up the genetic pathway that directs testis-cord formation. Smith *et al.*¹ show that knocking down *DMRT1* reduces the expression of the *SOX9* gene, a transcriptional activator whose male-specific upregulation is an early event in vertebrate sex determination, and which



Male factor: Smith and colleagues' work¹ centred on chicken embryos.

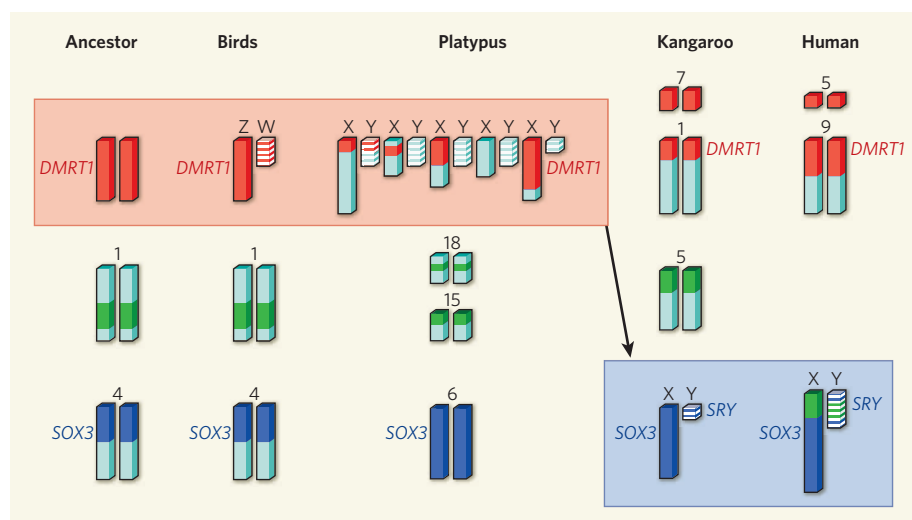


Figure 1 | Evolution of vertebrate sex chromosomes and sex-determining genes. The bird Z chromosome, but not the W, carries *DMRT1*, the gene shown by Smith *et al.*¹ to determine bird sex. The ZW sex chromosomes evolved from a pair of non-sex chromosomes, or autosomes, in an ancestral amniote (a vertebrate whose egg has an amniotic membrane) as the W degraded and left ZZ males and ZW females with different dosages of the *DMRT1* gene product. The platypus XY sex-chromosome complex shares genes with the bird ZW system (red), including *DMRT1*. In marsupial and placental mammals, the same ancestral Z chromosome region that includes *DMRT1* is present on autosomes (1 and 7 in kangaroo, 9 and 5 in human). The mammalian XY chromosomes were built up from two blocks (blue and green) that were autosomal in an ancestral amniote and remain autosomal in birds (chicken chromosomes 1 and 4) and platypus (chromosomes 6, 15 and 18). The blue region became a sex chromosome when *SOX3* mutated into the male-dominant *SRY*, and the green block was added later in the placental lineage. Thus the sex-determining function passed from an ancient Z chromosome bearing *DMRT1* (red shaded area) in an ancient amniote (and which still determines sex in birds, some lizards and possibly egg-laying mammals) to an autosome bearing *SOX3/SRY* (blue shaded area) in therian mammals (placentals and marsupials).

is regulated by *SRY* in mammals. In humans and mice, *DMRT1* is still dosage sensitive, but it seems to operate further downstream, being expressed in the testis after *SOX9* expression⁸.

The demonstration that *DMRT1* is the bird sex-determining gene closes a large and awkward gap in our understanding of the organization and evolution of vertebrate sex chromosomes. The case has been made⁹ that *DMRT1* has an ancient role in vertebrate sex determination. Among

reptiles with a wide range of sex-determining systems (including temperature-determined sex, and male Y-chromosome-determined and female W-chromosome-determined sex) there is at least one lizard species with a Z chromosome that contains the same suite of genes as the bird Z (complete with *DMRT1*)¹⁰, suggesting that *DMRT1* may have an ancient role in reptile sex determination. Even more extraordinary is the observation¹¹ that monotreme mammals

(the egg-laying platypus and the echidna, which diverged from all other mammals 166 million years ago at the base of the mammalian evolutionary tree) have a sex-chromosome complex that is unrelated to the mammal XY but shares genes with the bird ZW system, including *DMRT1*.

These two findings^{10,11} suggest that a bird-like ZW system was ancestral to all amniotes (birds, reptiles and mammals), and that it was only recently usurped by *SRY* in therian mammals (placentals and marsupials) (Fig. 1). The mammalian *SRY* gene is thought to have evolved from the conserved *SOX3* gene on the X chromosome¹², which is expressed in testis, brain and the central nervous system. This probably happened after *SOX3* on an ancestral autosome was truncated and fused with elements that enforced testis-specific expression.

The involvement of *DMRT1* with sex in such a variety of organisms leads us to ask if sex determination is more highly conserved than we thought? Or is *DMRT1* just a particularly handy gene that is independently used and reused for sex?

Jennifer A. Marshall Graves is at the Research School of Biology, Australian National University, Canberra, ACT 2601, Australia.
e-mail: jenny.graves@anu.edu.au

1. Smith, C. A. *et al.* *Nature* **461**, 267–271 (2009).
2. Sinclair, A. H. *et al.* *Nature* **346**, 240–244 (1990).
3. Smith, C. A., McClive, P. J., Western, P. S., Reed, K. J. & Sinclair, A. H. *Nature* **402**, 601–602 (1999).
4. Raymond, C. S., Murphy, M. W., O'Sullivan, M. G., Bardwell, V. J. & Zarkower, D. *Genes Dev.* **14**, 2587–2595 (2000).
5. Ferguson-Smith, M. *Sex. Dev.* **1**, 2–11 (2007).
6. Matsuda, M. *et al.* *Nature* **417**, 559–563 (2002).
7. Yoshimoto, S. *et al.* *Proc. Natl Acad. Sci. USA* **105**, 2469–2474 (2008).
8. Bagheri-Fam, S., Sinclair, A. H., Koopman, P. & Harley, V. R. *Int. J. Biochem. Cell Biol.* doi:10.1016/j.biocel.2009.07.001 (2009).
9. Graves, J. A. M. *Annu. Rev. Genet.* **42**, 565–586 (2008).
10. Kawai, A. *et al.* *Chromosoma* **118**, 43–51 (2009).
11. Veyrunes, F. *et al.* *Genome Res.* **18**, 965–973 (2008).
12. Foster, J. W. & Graves, J. A. M. *Proc. Natl Acad. Sci. USA* **91**, 1927–1931 (1994).

NANOTECHNOLOGY

A gentle jackhammer

Enrico Gnecco

A futuristic method of data storage depends on the 'write-read' action of a multitude of tiny silicon tips. The concept of dynamic superlubricity offers a way to avoid the wear that would otherwise cripple them.

Imagine a world without abrasive wear. Uncle Joe would not complain about the pain in his worn-out knee joint whenever he bends down to lace his shoes. You also wouldn't have to change the tyres of your car so often. On the other hand, polishing pastes would be ineffective in buffing up that beautiful silver tea service; and the traditional tools of the carpenter and manicurist wouldn't be of much use.

Such a crazy world does not exist, but it is what various researchers are working towards, at least at the nanoscale. Reporting in *Nature Nanotechnology*¹, one such group, Lantz *et al.* at IBM Research, describes the latest advance towards creating a wear-free state.

The context of their research is data storage, specifically in thermomechanical systems that involve the use of atomic force microscopes

(AFMs) and thin polymer films. Ultrahigh densities of data storage can be achieved in the form of holes in the polymer film made by an array of heated AFM probes. In the reverse 'read' procedure, the data can be retrieved by the probes. The silicon tips of the probes have to remain sharp for optimum system performance, but the trouble is that they don't. Hence the search for ways to extend tip lifetime, one of which is to reduce the wear to which tips are exposed.

The startling news from Lantz and co-workers¹ is that they have dragged a silicon tip across a polymer surface for the best part of a kilometre without damaging either the tip or the surface. Although the impressiveness of that performance may not at first be obvious, consider that the distance travelled and the tip size differ by eleven orders of magnitude — the equivalent of someone climbing a ladder from

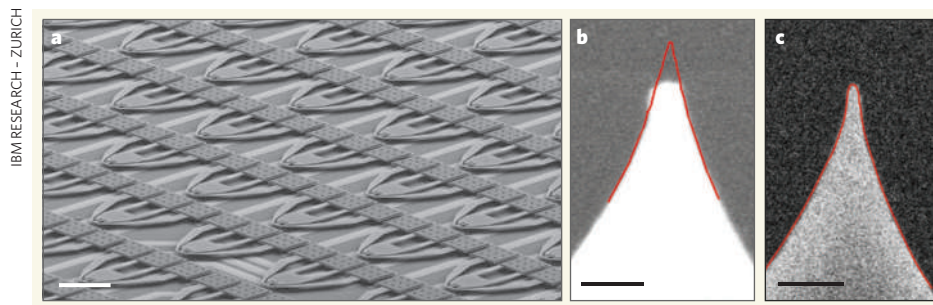


Figure 1 | Wears the tip? Millipede data-storage systems use arrays of atomic force microscopes (AFMs) to read and write data. The information is encoded as patterns of holes in a polymer film, which are recognized by arrays of AFM tips when they are dragged across the film's surface. **a**, Arrays of cantilevers (shown) support the tips, which sit at the apex of each cantilever. The tips are too small to be seen in this image. **b**, A problem with Millipede is that the tips wear out with use. In this close-up image of a used tip, the outline of the fresh tip is shown in red. **c**, Lantz *et al.*¹ report a method that eliminates wear by reducing the friction between the tip and the surface. This used tip therefore has the same outline as the fresh tip. (Scale bars: **a**, 40 μm ; **b**, **c**, 100 nm.)

Earth to Mars without wearing out the soles of their shoes. From their measurements, Lantz *et al.* estimate that the wear rate in their experiment was seven orders of magnitude lower than that usually observed on oil-lubricated steel (or, if you prefer, five orders lower than on diamond-like carbon film coatings, which are among the best solid lubricants available so far)². This result is indeed quite something.

But how was it achieved? To answer that, we have to go back to a paper by Socoliuc *et al.*³, which showed that atomic-scale friction is dramatically reduced when the contact between a sharp tip and a solid surface is modulated with superimposed vibrations at well-defined frequencies. While sliding on the surface, the tip has to cross a series of energy barriers, which results in abrupt jumps alternating with short periods of relaxation into minima of the surface potential (so-called stick-slip⁴). Crossing a barrier has a certain cost, and it can lead to irreversible damage to a tip if the normal load is not accurately controlled. Manipulating the contact region by inducing vibrations to it solves the problem — the amplitude of the energy barriers is periodically lowered, thus allowing motion with negligible friction. This concept, 'dynamic superlubricity', has quickly found applications in scanning probe microscopy, and several crystal lattices have now been gently resolved by AFMs that have vibrating contact tips⁵.

Lantz *et al.*¹ have adopted dynamic superlubricity and taken it out of the surface-physics lab. The tips that they used are parts of the elaborate thermal probes that can locally heat polymer surfaces above the glass-transition temperature (usually around 400 °C), producing tiny holes in the polymer. Each hole can be subsequently located with the same tip and interpreted as a bit of information. Using arrays with thousands of nanoprobe, data rates of the order of a few megabytes per second can be achieved, with areal densities of 150 gigabits per square centimetre and very low power consumption⁶.

A practical application of these arrays as efficient data-storage devices is being developed

by IBM, and goes under the name of Millipede, because of its multitude of 'legs' in the form of very fine silicon tips (Fig. 1). But here we come back to the wear issue. Lantz *et al.*¹ estimate that, under the best-imaginable conditions, the wear rate affecting a Millipede device can still be four orders of magnitude higher than the maximum tolerable value for proper functioning. This won't be the case if, as the authors have shown, the Millipede legs are periodically vibrated while 'walking' by applying an a.c. voltage between the surface and a platform close to the probing tip. The result is that the vibrating tip moves smoothly across the surface, without losing contact, and without losing the possibility of heating the polymer to produce a hole. Their success in using this gentle jackhammer is exemplified by scanning-

electron-microscope images of tips before and after sliding a distance of 750 m, with and without modulation at 500 kHz (Fig. 1).

The choice of the excitation frequency (and of the corresponding amplitude of a few volts) stems from the authors' analysis of 'frictional spectra', normal and lateral resonance curves, finite-element simulations and friction-versus-amplitude curves, all of which complement the main result of the work. The mechanism of wear reduction works best when the modulation frequency is close to the normal resonance of the system formed by the tip mechanically coupled to the surface, as Socoliuc *et al.*³ noted in the context of friction reduction.

Exploitation of the principle of dynamic superlubricity could feasibly be extended to other micro- and nanoelectromechanical systems that have tiny contacting parts and, more generally, to lithography and manufacturing processes carried out on small scales. In the long term, it is not beyond the realms of possibility that even Uncle Joe's knee will benefit from this effect. ■

Enrico Gnecco is in the National Center of Competence in Research 'Nanoscale Science', Department of Physics, University of Basel, Klingelbergstrasse 82, CH-4056 Basel, Switzerland.
e-mail: enrico.gnecco@unibas.ch

1. Lantz, M. A., Wiesmann, D. & Gotsmann, B. *Nature Nanotech.* doi:10.1038/nnano.2009.199 (2009).
2. Erdemir, A. & Martin, J.-M. (eds) *Superlubricity* (Elsevier, 2007).
3. Socoliuc, A. *et al. Science* **313**, 207–210 (2006).
4. Schirmeisen, A. *et al. Phys. Rev. B* **71**, 245403 (2005).
5. Gnecco, E. *et al. Nanotechnology* **20**, 025501 (2009).
6. Knoll, A. *et al. Microelec. Eng.* **83**, 1692–1697 (2006).

EARLY EARTH

Oxygen for heavy-metal fans

Timothy W. Lyons and Christopher T. Reinhard

Chromium isotopes provide an eyebrow-raising history of oxygenation of Earth's atmosphere. Not least, it seems that oxygen might have all but disappeared half a billion years after its initial rise.

That the Earth's early atmosphere had vanishingly low levels of oxygen is widely, if not universally¹, acknowledged. The first appreciable rise occurred about 2.4 billion years ago, during what is known as the Great Oxidation Event, or GOE. Evidence for a second major increase about 750 million years ago — this one large enough to favour the rise of animals — has received an equally enthusiastic reception. The details of these increases, however, and the nature of the ocean and atmosphere during the time between them, are matters of debate. Frei *et al.*² (page 250 of this issue) have used a new geochemical tracer, chromium (Cr) isotopes, to reveal that oxygenation was

probably more nuanced and less unidirectional than we might have imagined.

Before the GOE, oxygen levels in the atmosphere were clearly much less than 1% of today's value. Of the many lines of evidence pointing to an oxygen-poor atmosphere, most convincing are diagnostic sulphur-isotope fingerprints recorded in minerals such as pyrite (FeS_2)^{3,4}. Geochemists have long considered the abundant iron formations that accumulated on the sea floor to be key evidence that the ocean beneath this early atmosphere was similarly oxygen-lean and frequently rich in dissolved iron (Fig. 1, overleaf). Yet the relationship between the subsequent rise in atmospheric



50 YEARS AGO

My Philosophical Development.

By Bertrand Russell — All those whose study of philosophy is grounded in the empirical tradition regard Lord Russell as the greatest living philosopher ... Although one should not neglect other influences ... there is no doubt that the main responsibility for the present state of philosophy lies squarely on Russell's shoulders ... There are few philosophers in history who have written important philosophical works almost continuously for fifty years: Russell has added to the immense debt we owe him by now giving us a full-scale account of his philosophical development, written with all the clarity, verve and wit we are accustomed to expect from anything he writes.

From *Nature* 12 September 1959.

100 YEARS AGO

Organic Memory. By Prof.

Richard Semon — The theory of the Mneme, propounded by Prof. Semon, has attracted the attention both of psychologists and of those naturalists who are interested in the profound problems of hereditary transmission. It is founded on the statement, which everyone is ready to admit, that a stimulus must affect the quality of living matter in such a way that the matter is not the same as it was before the stimulus acted. A permanent change, which, in a sense, may be called a memory, has been effected, or, to use the terminology invented by Semon, the action has been engraphic and the change itself is an engram ... All stimuli then produce engrams, and the sum of the engrams of a living being is its mneme ... Thus a stimulus may produce effects which radiate from the organised matter first affected to organised matter throughout the whole organism, either by nerve paths or by proplasmic intercellular filaments, and in this way faint engrams may be made on the matter of the reproductive elements, ova or spermatozooids.

From *Nature* 9 September 1909.

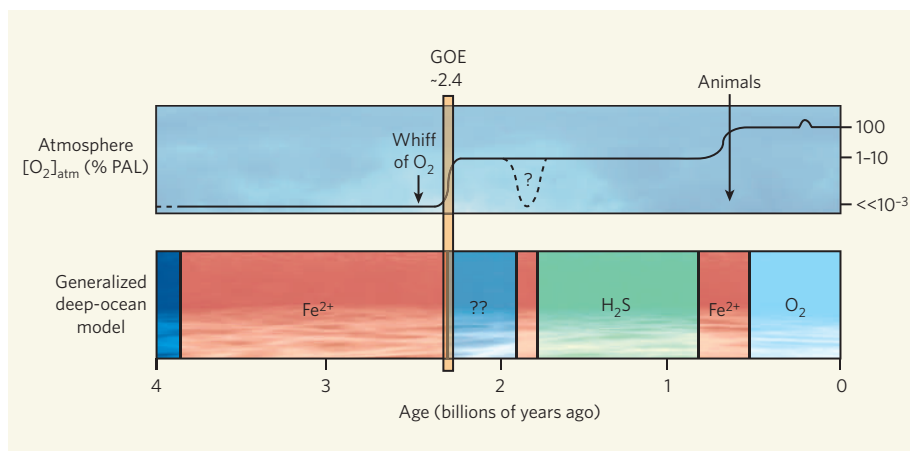


Figure 1 | Tracking oxygen above and below. A rise in atmospheric oxygen about 2.4 billion years ago at the Great Oxidation Event (GOE) coincided with the end of frequent and protracted, iron (Fe^{2+})-rich, oxygen-poor conditions in the deep ocean. Several studies^{2,6,7}, however, point to at least transient oxygen increases well before the GOE — evidence of early, oxygen-producing photosynthesis⁵. Through their analysis of chromium-isotope ratios in iron formations, Frei *et al.*² argue that the return of an iron ocean, peaking about 1.9 billion years ago, was triggered by a dramatic decline in oxygen, perhaps to values approaching those in the atmosphere before the GOE. Next, a billion years or more of ubiquitous hydrogen sulphide (H_2S) in the deep ocean may have resulted from a rise in atmospheric oxygen^{2,10–12}. A major step in oxygenation then followed at the dawn of animal life¹³. Oxygen concentrations are given in per cent of the present atmospheric level (PAL). (Figure modified from refs 15 and 16.)

oxygen and the parallel chemical trends in the deep ocean remains enigmatic. In particular, researchers have been hard-pressed to explain a surprising return of iron formations half a billion years after the GOE.

Difficult questions require clever solutions, and Frei *et al.*² have sought insight from the iron formations themselves. Deposited with the iron were stable isotopes of chromium, which the authors interpret as a tracer for oxygen content in the atmosphere. Chromium is immobile in continental crust beneath an anoxic atmosphere and happily resides in its reduced or +3 state. But with increasing atmospheric oxygen comes the oxidation of manganese (Mn), another metal in the crust and overlying soil, and the resulting MnO_2 is able to react with chromium, oxidizing it in the process. The oxidized chromium loses electrons, with a shift in redox state or valence from +3 to +6, and it also gains mobility — meaning that rainwater can remove it from the soil, and rivers can transport it to the ocean. This mobilized chromium differs in one other notable way: during reaction with MnO_2 , the heavy isotope ^{53}Cr is oxidized preferentially relative to the lighter ^{52}Cr , and this preference, too, is transferred to the ocean.

In the ocean, the ^{53}Cr -enriched chromium is once again immobilized as it is re-reduced to the +3 form through reaction with iron and is deposited within iron formations. Because of the high efficiency of this reaction, it is assumed that all the chromium is stripped from the sea water, minimizing any concerns about isotopic discrimination at this step. In other words, the isotopic properties of the initial products of weathering can be transferred to the ocean, and are captured and preserved unaltered over

geological time. Measuring the chromium-isotope properties in iron formations of different ages is thus a window on to the evolving oxygenation state of the atmosphere.

Frei *et al.*² took a long look through this window, and found a few surprises. One of them is evidence for elevated but still low oxygen in the atmosphere long before the GOE. Enrichments of ^{53}Cr in iron formations all but confirm the biological production of oxygen by photosynthesis at least 300 million years before the GOE. This assertion is certain to please those who have come to the same conclusion by very different paths^{5–7} and displease others^{8,9} who consider that the first appreciable accumulation of oxygen in the atmosphere at the GOE marked its first production.

But no matter where one falls on the timing issue, most people probably imagine that once oxygen really started accumulating in the atmosphere, it never looked back. Even if you allow for a bit of a roller-coaster ride, that's nothing like what Frei *et al.* claim: a reprise of near pre-GOE values roughly 500 million years later, perhaps ushering in the return of an iron-rich ocean as indicated by a second big wave of iron formations (Fig. 1). The chromium isotopes suggest that soon after that there was another rise in atmospheric oxygen, this one large enough to have triggered more than a billion years of pervasive hydrogen sulphide accumulation in the deep ocean^{10,11}. The thinking is that pyrite weathering on the continents in the presence of increasing atmospheric oxygen led to greater sulphate delivery to the still-oxygen-poor deep ocean, where bacteria re-reduced the sulphate to hydrogen sulphide (H_2S). Such conditions, rare today, could have set the course for the early evolution of

eukaryotic life^{11,12}. The story culminates in the largest observed shift in the chromium-isotope data, about 750 million years ago, corresponding to a big leap in atmospheric oxygen to levels that welcomed the first animals¹³.

So far, so promising. But even though this is still the tracer's honeymoon period, we already have to wonder whether the levels of early oxygen needed to generate the required MnO₂ for chromium oxidation, and thus ⁵³Cr enrichment, are beyond the upper limits set by the sulphur isotopes before the GOE. If they are, the authors have a problem.

The greatest revelation of this study², the plunge to early pre-GOE chromium-isotope values 500 million years after the GOE, may also be its weak point. By all previous accounts, such a dramatic change in atmospheric conditions should be recorded in the diagnostic sulphur-isotope fingerprint for low oxygen. But it is not¹⁴. Frei *et al.* may simply be taking the isotope values too literally in assuming that

variations in chromium isotopes scale in a simple way to varying atmospheric oxygen. They freely confess that the sensitivity of the isotopic systematics to atmospheric oxygen may not be a simple linear relationship, and would probably agree that we still have a lot to learn about chromium cycling on land and in the ocean.

The story of chromium isotopes as an oxygen tracer is not perfect. For now, though, it provides one of the more satisfying explanations for the surprising return of voluminous iron formations some half a billion years after the GOE, and it provides legitimate fodder for the debate over the onset of biological oxygen production. Regardless of how the approach matures, we are sure to be left with some tantalizing possibilities and, even better, a new tool for refining our view of the early ocean. That's not bad for an element best known for putting the shine on car bumpers.

Timothy W. Lyons and Christopher T. Reinhard are in the Department of Earth Sciences,

University of California, Riverside,
California 92521, USA.
e-mail: timothy.lyons@ucr.edu

1. Ohmoto, H., Watanabe, Y., Ikemi, H., Poulson, S. R. & Taylor, B. E. *Nature* **442**, 908–911 (2006).
2. Frei, R., Gaucher, C., Poulton, S. W. & Canfield, D. E. *Nature* **461**, 250–253 (2009).
3. Farquhar, J. & Wing, B. A. *Earth Planet. Sci. Lett.* **213**, 1–13 (2003).
4. Bekker, A. *et al.* *Nature* **427**, 117–120 (2004).
5. Brocks, J. J., Logan, G. A., Buick, R. & Summons, R. E. *Science* **285**, 1033–1036 (1999).
6. Anbar, A. D. *et al.* *Science* **317**, 1903–1906 (2007).
7. Kump, L. R. & Barley, M. E. *Nature* **448**, 1033–1036 (2007).
8. Kopp, R. E., Kirschvink, J. L., Hilburn, I. A. & Nash, C. Z. *Proc. Natl Acad. Sci. USA* **102**, 11131–11136 (2005).
9. Rasmussen, B., Fletcher, J. R., Brocks, J. J. & Kilburn, M. R. *Nature* **455**, 1101–1104 (2008).
10. Canfield, D. E. *Nature* **396**, 450–453 (1998).
11. Lyons, T. W., Anbar, A. D., Severmann, S., Scott, C. & Gill, B. C. *Annu. Rev. Earth Planet. Sci.* **37**, 507–534 (2009).
12. Anbar, A. D. & Knoll, A. H. *Science* **297**, 1137–1142 (2002).
13. Love, G. D. *et al.* *Nature* **457**, 718–721 (2009).
14. Johnston, D. T. *et al.* *Geochim. Cosmochim. Acta* **70**, 5723–5739 (2006).
15. Canfield, D. E. *Annu. Rev. Earth Planet. Sci.* **33**, 1–36 (2005).
16. Kump, L. R. *Nature* **451**, 277–278 (2008).

CELL BIOLOGY

Sent by the scent of death

Christopher Gregory

Dying cells release 'find-me' factors that attract professional scavenger cells to engulf and digest them. These cellular invitations to dine can take unexpected forms.

About one million cells die every second in our bodies, often by the process of programmed cell death, or apoptosis. Dead cells are either expelled to the outside — for instance, by sloughing off of effete skin or gut cells — or are rapidly engulfed by healthy neighbours or professional scavenger cells. Notable among the scavengers are the macrophages — 'big eating' white blood cells that are derived from blood monocytes. It has long been assumed that apoptotic cells release compounds that recruit monocytes and macrophages to sites of cell death, but the full identity of these factors is not known. In this issue (page 282), Elliott *et al.*¹ reveal one group of 'find-me' signals as the nucleotides ATP and UTP — better known as providers of metabolic energy, mediators of intracellular signalling and, in alternative forms, as components of DNA and RNA. The authors show that macrophages detect nucleotides using specific receptors, a mechanism that ensures their rapid movement towards apoptotic cells, which is a necessary prelude to engulfment.

Elliott and colleagues¹ analysed the movement of isolated monocytes and macrophages, both in mammalian cell culture and in mice. In the latter, they used an air-pouch model — a type of skin blister that allows the study of directed movement of white blood cells *in vivo*. They showed that cells

undergoing apoptosis release soluble factors that selectively attract monocytes and macrophages. Release of these factors is directly linked to the induction of apoptosis and is not due to passive leakage of cellular content across a damaged cell membrane. By contrast, injection of bacterial products into the pouch results in the recruitment of neutrophils — a different type of professional scavenger cell that is usually mobilized to sites of damage

and infection to engulf microorganisms.

The compounds released from apoptotic cells lose their attractant properties when they are exposed to enzymes that degrade nucleotides¹, providing the first clue to their molecular identities. The authors followed up on this lead to confirm that it is indeed nucleotides, specifically ATP and UTP, that are released from apoptotic cells. They showed that ATP and UTP generate directional migration (chemotaxis) in monocytes and macrophages, and are optimally active at concentrations similar to those found in supernatants of apoptotic cells¹.

Nucleotides are known to bind to P2Y receptors — members of the G-protein-coupled receptor (GPCR) family. GPCRs bind to different types of molecule to signal responses in many cell types. Other GPCRs on macrophages are already known to transduce signals for chemotaxis towards apoptotic cells that release fractalkine² (CX₃CL1) or the lipid

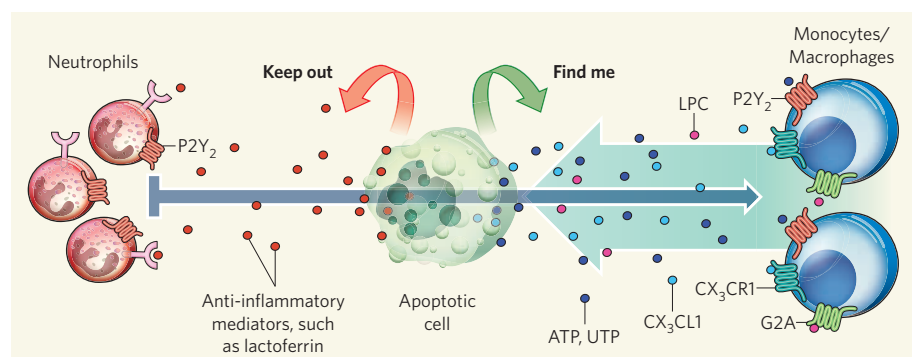


Figure 1 | Balancing act. Elliott and colleagues¹ show that the nucleotides ATP and UTP, released by apoptotic cells, function as 'find-me' signals that bind to the G-protein-coupled receptor (GPCR) P2Y₂ on macrophages and monocytes. Other chemoattractants released from apoptotic cells, notably the chemokine fractalkine (CX₃CL1) and the lipid lysophosphatidylcholine (LPC), also bind to GPCRs (CX₃CR1 and G2A, respectively). Because neutrophils also express P2Y receptors and migrate in response to nucleotides as part of the inflammatory response, selective movement of monocytes and macrophages to sites of cell death could be mediated through the balance between 'find-me' signals and anti-inflammatory 'keep-out' signals (for example, lactoferrin) that are released by apoptotic cells.

chemoattractant lysophosphatidylcholine³ (Fig. 1). Elliott *et al.*¹ used various approaches to show that the expression by monocytes and macrophages of a specific P2Y receptor, P2Y₂, is required, at least in part, for ATP- and UTP-mediated migration.

To study the significance of the P2Y₂–nucleotide interaction for efficient clearance of apoptotic cells in a physiological setting, the authors¹ shifted their attention to the mouse thymus, in which massive and synchronous apoptosis can be induced with a steroid hormone. They showed that adding either an enzyme that breaks down ATP and UTP, or a P2Y₂ inhibitor, compromises the macrophage-mediated clearance process that usually accompanies thymic apoptosis. In addition, apoptotic cells persist in the thymus of P2Y₂-deficient mice. So it seems that nucleotide-mediated signalling is necessary for the efficient engulfment of apoptotic cells, at least in the thymus.

Exciting new findings always generate a wealth of questions, and the work of Elliott

et al. is no exception. Efficient clearance of apoptotic cells is known to be important for preventing autoimmune disease⁴ and, given this association, it will be interesting to determine whether P2Y₂-deficient mice show signs of autoimmunity. It will also be instructive to find out how the active release of nucleotides is coupled to the biochemical events that control apoptosis. Is nucleotide release a feature of all dying cells? How important is this process for engulfment of dying cells by non-professional phagocytes or for other professional phagocytes, such as dendritic cells? As several other chemoattractants are released from apoptotic cells (Fig. 1), what is the relative importance of nucleotide release in different tissue settings?

Intriguingly, the nucleotide ATP has been shown⁵ to activate migratory activity in neutrophils, which, despite their specialist scavenging functions — part of the inflammatory response — are not the usual 'clean-up' operatives at sites of apoptosis. As apoptotic cells produce anti-inflammatory mediators, including

suppressors of neutrophil migration⁶, selective attraction of monocytes and macrophages by nucleotides released from apoptotic cells may depend not only on the 'find-me' action of nucleotides on monocytes and macrophages, but also on the 'keep-out' action of factors, such as lactoferrin⁶, that act on pro-inflammatory scavengers (Fig. 1). The burning questions are whether this nucleotide–receptor-mediated chemotaxis can be targeted either to promote apoptotic-cell clearance to treat inflammatory or autoimmune disorders, or perhaps to inhibit clearance of apoptotic cells to stimulate inflammatory or immune responses against tumours.

Christopher Gregory is at the Centre for Inflammation Research, University of Edinburgh and ImmunoSolv Limited, Edinburgh EH16 4SB, UK. e-mail: chris.gregory@ed.ac.uk

1. Elliott, M. R. *et al.* *Nature* **461**, 282–286 (2009).
2. Truman, L. A. *et al.* *Blood* **112**, 5026–5036 (2008).
3. Peter, C. *et al.* *J. Biol. Chem.* **283**, 5296–5305 (2008).
4. Nagata, S. *Immunol. Rev.* **220**, 237–250 (2007).
5. Chen, Y. *et al.* *Science* **314**, 1792–1795 (2006).
6. Bournazou, I. *et al.* *J. Clin. Invest.* **119**, 20–32 (2009).

MATERIALS CHEMISTRY

Catalysts made thinner

Avelino Corma

Thinner can be better, at least for the industrially useful catalysts known as zeolites. A technique that allows single layers of zeolites to assemble from solution opens up a plethora of practical applications.

The synthesis of solid, crystalline materials that have ordered arrays of pores has long been an attractive goal. Such materials were first used as molecular sieves or as 'sponges' to store other compounds. But by fine-tuning the pore dimensions and the composition of the crystal network, they could also be used in many other applications that require specific molecules to be recognized and bound — such as in catalysis, in substrate-selective membranes, or in chemical sensors, to name but a few¹. For some of these applications, thin layers of porous materials are highly desirable, but no method was available by which such samples could be directly prepared. On page 246 of this issue, Choi *et al.*² describe just such a method for synthesizing exceptionally thin sheets of porous materials known as zeolites.

Zeolites are crystalline, inorganic solids formed from interlocked, tetrahedral TO₄ units (where T can be one of many different ions, such as Si⁴⁺, P⁵⁺, Al³⁺ and so on). Each oxygen is shared between adjacent tetrahedra, so that the overall molecular formula of the framework is TO₂. The compounds are normally prepared in hot water under pressure, in many cases using organic cations as structure-directing agents (SDAs) that guide the assembly of the inorganic components. Typical procedures

involve mixing a source of T atoms, a mineralizing agent (hydroxide and/or fluoride ions) and the SDA in water at temperatures between about 100 and 200 °C. The resulting materials have cavities that are filled by SDAs, which are then removed, leaving behind empty pores. The self-assembly of zeolites using an organic SDA is thus a complex process that involves numerous weak interactions of the reaction components with the SDA. It seems that SDAs directly influence pore topology by closely matching the geometrical arrangement of the inorganic framework.

Until recently, zeolites had pore dimensions

of 1 nanometre or less. This limits the size of molecules that can be stored or reacted within zeolite pores, and has spurred researchers to find ways of making materials that have larger cavities. Much of this work has focused on using organic SDAs of different dimensions to direct the synthesis towards making certain secondary building units^{3,4}. In this way, a three-dimensional zeolite has been prepared that has extra-large pores of about 1.2 nanometres in diameter⁵, presenting interesting opportunities for catalysis. More recently, the synthesis of the first mesoporous zeolite (which contains pores of about 2 nanometres in diameter) was reported⁶.

Other efforts to improve the functional properties of zeolites have focused on finding ways to make new zeolite morphologies, such as nanocrystals⁷, multicrystal arrays⁸, amorphous mesoporous materials⁹ and many more. Of these, single layers of zeolites that have high surface areas are of great interest for several catalytic applications^{10,11} — for example, when reacting large molecules that cannot readily access the interiors of bulk zeolites; in catalysis

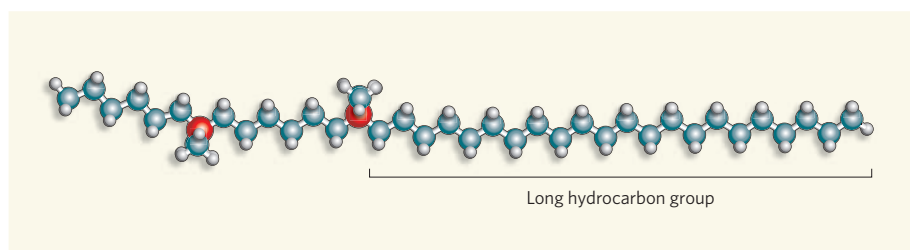


Figure 1 | Structure-directing agent for a two-dimensional zeolite. Choi *et al.*² used this organic molecule as a structure-directing agent (SDA) to control the assembly of nanosheets of a zeolite — a porous, crystalline, inorganic material known as ZSM-5. Two positively charged nitrogen atoms (red) in the SDA control the assembly of the inorganic framework, while a long hydrocarbon group prevents growth of the zeolite in the direction in which that group extends.

that occurs at pore mouths, rather than in pore interiors; and to improve catalytic processes where the internal pores of three-dimensional zeolites form undesired side products or deactivate the catalyst.

Single layers of zeolites have previously been prepared by the post-synthetic treatment of three-dimensional structures, and have proven benefits as catalysts^{10,11}. Choi *et al.*² now report the first direct synthetic method for preparing zeolite nanosheets — in this case, for a compound known as ZSM-5 aluminosilicate, one of the most important catalysts used in the petrochemicals industry. The trick is to use an organic molecule that has a long hydrocarbon chain and two cationic groups that are separated by a shorter linkage (Fig. 1). The cationic groups act as SDAs to crystallize the ZSM-5 structure, whereas the long hydrophobic chain limits the growth of the zeolite crystal along that direction.

The authors used electron microscopy to confirm that, using their SDA, they obtained a layer of ZSM-5 that was just 2 nanometres thick. This corresponds to the thickness of a single unit cell (the smallest subunit of the crystal structure). On heating, the SDA vacates the pores, leaving a unilamellar zeolite that has a surface area of roughly $700 \text{ m}^2 \text{ g}^{-1}$ — almost twice that of the conventional form of the zeolite.

Choi *et al.* report that their prepared material has strong acidity and high thermal and hydrothermal stability — for example, it maintains a sizeable fraction of the aluminium in the zeolite framework even after being heated in steam at 700°C . These properties are crucial for catalytic applications. Indeed, the authors found that the catalytic activity of their nanosheets in several reactions of industrial interest is much higher than for typical ZSM-5.

The authors also found that their unilamellar ZSM-5 was a much longer-lived catalyst than conventional forms of the zeolite in a process that transforms methanol into petrol. In this process, the catalyst deactivates because coke — a carbon side-product of the process — builds up in the pores. Choi *et al.* found that coke forms mostly on the external surface of their nanosheets, rather than inside the pores, as occurs in normal ZSM-5. They propose that the observed external build-up of coke poses less of a barrier to reactant molecules trying to diffuse into the catalytic pores of the zeolite than the accumulation of coke inside pores, and that this explains the longer catalytic lifetime of ZSM-5 nanosheets.

I believe that Choi and colleagues' direct synthesis of zeolite layers will allow improved catalysts to be made for processes that require pore-mouth catalysis. The nanosheets are useful for processing large molecules in oil refining and petrochemistry, as well as for the production of bulk and fine chemicals. I also predict that unilamellar and multilamellar ZSM-5 materials will be excellent catalysts (or co-catalysts) for 'cracking' processes

that convert large hydrocarbon molecules into petrol and diesel, or that produce large yields of propylene (a valuable feedstock used by the chemical industry).

Moreover, Choi and colleagues' synthetic strategy for making nanosheets could be applied to other zeolites that have pores running along only one direction. Such materials would be excellent for preparing selective membranes that allow only certain molecules through. In fact, intense research efforts have focused on reducing the thickness of membranes in order to increase permeability, while maintaining a preferred orientation of pores, for separation applications. Zeolite nanosheets might be the solution to this problem. Finally, catalytic transition-metal complexes could also be grafted to the surfaces of zeolite sheets to generate multifunctional catalysts¹². Choi and colleagues' work thus certainly enlarges the number of possibilities for zeolites in future applications. ■

Avelino Corma is at the Instituto de Tecnología Química (UPV-CSIC), Universidad Politécnica de Valencia, Avenida Los Naranjos, Valencia, Valencia 46022, Spain.

e-mail: acorma@itq.upv.es

1. Davis, M. E. *Nature* **417**, 813–821 (2002).
2. Choi, M. *et al.* *Nature* **461**, 246–249 (2009).
3. Cheetham, A. K. *et al.* *Stud. Surf. Sci. Catal.* **135**, 788–795 (2001).
4. Corma, A., Navarro, M. T., Rey, F., Rius, J. & Valencia, S. *Angew. Chem. Int. Edn* **40**, 2277–2280 (2001).
5. Corma, A., Diaz-Cabañas, M. J., Jorda, J. L., Martinez, C. & Moliner, M. *Nature* **443**, 842–845 (2006).
6. Sun, J. *et al.* *Nature* **458**, 1154–1157 (2009).
7. Lovallo, M. C. & Tsapatsis, M. in *Advanced Techniques in Catalyst Preparation* (ed. Moser, W. R.) 307–343 (Academic, 1996).
8. Lee, J. S., Lee, Y. J., Tae, E. L., Park, Y. S. & Yoon, K. B. *Science* **301**, 818–821 (2003).
9. Beck, J. S. *et al.* *J. Am. Chem. Soc.* **114**, 10834–10843 (1992).
10. Corma, A., Fornes, V., Pergher, S. B., Maesen, T. L. M. & Buglass, J. G. *Nature* **396**, 353–356 (1998).
11. Corma, A., Botella, P. & Mitchell, C. *Chem. Commun.* 2008–2010 (2004).
12. Corma, A., González-Arellano, C., Iglesias, M., Navarro, M. T. & Sánchez, F. *Chem. Commun.* 6218–6220 (2008).

DEVELOPMENTAL BIOLOGY

Instructions writ in blood

Tariq Enver and Sten Eirik W. Jacobsen

It seems that growth factors may instruct blood-cell progenitors to develop into specific mature cell types, actively determining lineage choice. But is this reductionist view of cell fate overly simplistic?

The nature-versus-nurture debate about human potential will be familiar to most. A similar dispute surrounds the role of environmental signals, such as growth factors, in determining cell-type identity in multicellular organisms. Two papers, one by Rieger *et al.*¹ published in *Science*, and the other by Sarrazin *et al.*² in *Cell*, inform this discussion by providing evidence that environmental cues might actively determine cell lineage.

The process of lineage specification is fundamental to the development and maintenance of tissues in multicellular organisms. Blood formation, or haematopoiesis, exemplifies this. Here, rare multipotent stem cells and progenitor cells that reside in the bone marrow produce several different lineages of mature blood cell. In steady-state conditions, haematopoiesis accommodates a ferociously high rate of cell turnover, with millions of cells being generated every second. In addition, blood cells are produced at speed 'on demand' in response to injury or challenge — for instance, red blood cells in response to bleeding or white blood cells in response to infection.

With a myriad of defined cellular intermediates along any lineage pathway^{3,4}, and a battery of functional assays enabling studies of the generation of different blood-cell lineages from single stem cells and progenitor cells^{3–5}, the blood system has for decades provided

a playground for exploring mechanisms of lineage specification. From this work an impressive list of nuclear, cytoplasmic and extracellular regulators of blood-cell development have been identified and characterized^{3,4}, of which some growth factors (termed cytokines) are routinely used in the clinic to replenish specific types of blood cell⁵. These growth factors function by selectively enhancing the proliferation and differentiation of progenitors that are already restricted in their lineage options, and by improving the function of their mature descendants. However, the patterns of expression of growth-factor receptors on multipotent blood-cell progenitors, and their responsiveness to these factors⁵, suggest that growth factors might also target multipotent progenitors or stem cells at a pre-committed stage, possibly by affecting their lineage choice per se.

The unresolved question therefore is whether environmental or extrinsic cues such as growth factors usually instruct multipotent progenitors to commit to a particular lineage — the 'instructive' model (Fig. 1a, overleaf); or whether these cues simply support the survival and proliferation of cells that have already committed to a specific lineage by a cell-autonomous or intrinsic agency — the 'selective' or 'permissive' model (Fig. 1b). This question is by no means new, and evidence has been provided

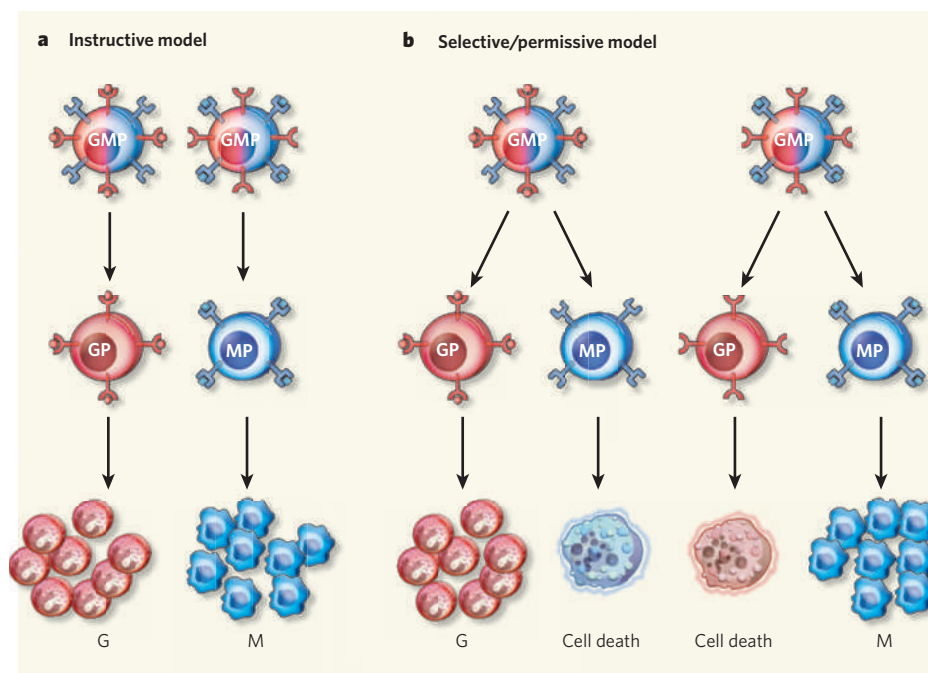


Figure 1 | Instructive and selective models of lineage choice. Receptors for the blood-cell growth factors granulocyte colony-stimulating factor (G-CSF, red) and macrophage colony-stimulating factor (M-CSF, blue) are expressed on bipotent granulocyte-macrophage progenitors (GMP) and on multipotent stem cells (not shown), as well as on lineage-restricted granulocyte progenitors (GP) or macrophage progenitors (MP). **a**, In the instructive model, supported by data from Rieger *et al.*¹ and Sarrazin *et al.*², growth factors bind to their receptors and instruct progenitor cells to commit to a particular lineage — granulocyte (G) or macrophage (M). **b**, In the selective or permissive model, a cell-autonomous process drives commitment of blood cells to a distinct cell lineage. Here, growth factors are needed to sustain the survival of committed progenitor cells. The outcome — a pool of specific mature cell types — is identical in both cases.

(and debated) in support of permissive^{5–7} and instructive^{5,8,9} actions of cytokines in determining blood-cell-lineage commitment.

Ectopic-expression experiments have demonstrated the instructive capacity of cytokines^{8,9}, but the artificial nature of this approach — the genes encoding cytokine receptors are introduced into cells — leaves uncertainty about the biological relevance of these findings. The more physiological approach of loss-of-function experiments — using knockout or knock-down of cytokines or their receptors^{3–5} — has yet to demonstrate that cytokines affect any blood-cell lineage at the stage of lineage commitment, as would be expected if they were acting through instructive mechanisms.

Part of the reason that this issue has been difficult to address is that the outcome measured in these studies — the production of lineage-specified cells — is the same whether regimes are instructive or selective (Fig. 1). To unequivocally distinguish between instructive and selective effects on lineage development, one must account for the fate, be it survival and proliferation, death or quiescence, of every daughter cell of an uncommitted progenitor cell that has more than one lineage potential.

Rieger *et al.*¹ approached this daunting task with the view that ‘seeing is believing’. They performed continuous single-cell imaging of bipotent granulocyte-macrophage progenitors carrying a fluorescent reporter of lineage

commitment; the progenitors were cultured with either granulocyte colony-stimulating factor (G-CSF) or macrophage colony-stimulating factor (M-CSF). These cytokines selectively promote the development of different types of white blood cell; G-CSF promotes development of granulocytes and M-CSF of macrophages. Both G-CSF and M-CSF have been proposed to function in an instructive manner^{10,11} in the lineage commitment of granulocyte-macrophage progenitors. Although only a fraction of the progenitors in Rieger and colleagues’ culture system¹ was bipotent at the outset, and a significant number of progenitor cells died or failed to grow, the authors provide cogent evidence that G-CSF and M-CSF can instruct lineage commitment.

However, the physiological relevance of these results remains uncertain. Rieger and colleagues’ experiments were carried out in artificial cell-culture systems using high concentrations of single growth factors, whereas decisive experiments will require faithful mapping of the fate of single cells in live animals — this is currently not possible in mammalian models such as mice. Those caveats aside, this work¹ elegantly demonstrates the importance and power of high-resolution imaging of cell fate at the single-cell level.

In a related paper, Sarrazin *et al.*² present findings implicating the interplay between transcription factors and growth-factor signalling

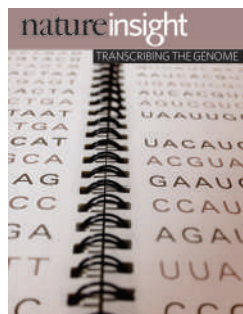
in decisions about blood-cell fate, in line with previous studies¹². They show that the transcription factor MafB selectively restricts the responsiveness of multipotent progenitors, and perhaps stem cells, to M-CSF. Crucially, studies in mice² reveal that the interplay between MafB and M-CSF limits the development of cell lineages that are usually promoted by M-CSF. The authors also provide data suggesting that MafB functions by inhibiting the instructive actions of M-CSF in stem cells. The experiments of Sarrazin and colleagues² are, however, not definitive on this point, as the markers used to identify multipotent stem cells and granulocyte-lineage commitment are equivocal.

So, although fascinating, should the ‘selective versus instructive’ and ‘intrinsic versus extrinsic’ commitment debate really be relentlessly pursued, especially if the available technology does not allow us to perform the definitive experiments? The debate started at a time when little was known about the molecular machinery that controls blood-cell-fate decisions. But this picture is changing as we begin to understand the nature and organization of the intrinsic transcriptional regulatory circuits associated with blood-cell-lineage specification¹³. The role of extrinsic, niche-associated cues or signals is also becoming increasingly clear¹⁴, although little is known about how these extrinsic and intrinsic regulators interact in networks^{3,4,12}. Perhaps the challenge is not to try to determine, in a reductionist manner, who is master and who is servant in these interrelated circuits, but rather to decipher the dynamics of the circuits that determine cell fate¹⁵. Precisely who is in the driving seat at any given juncture may be more a matter of context than rule.

Tariq Enver is in the MRC Molecular Haematology Unit, and Sten Eirik W. Jacobsen is in the Haematopoietic Stem Cell Laboratory, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, UK.

e-mails: tariq.enver@imm.ox.ac.uk; sten.jacobsen@imm.ox.ac.uk

1. Rieger, M. A., Hoppe, P. S., Smejkal, B. M., Eitelhuber, A. C. & Schroeder, T. *Science* **325**, 217–218 (2009).
2. Sarrazin, S. *et al.* *Cell* **138**, 300–313 (2009).
3. Laiosa, C. V., Stadtfeld, M. & Graf, T. *Annu. Rev. Immunol.* **24**, 705–738 (2006).
4. Orkin, S. H. & Zon, L. I. *Cell* **132**, 631–644 (2008).
5. Metcalf, D. *Blood* **111**, 485–491 (2008).
6. Fairbairn, L. J., Cowling, G. J., Reipert, B. M. & Dexter, T. M. *Cell* **74**, 823–832 (1993).
7. Stoffel, R. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 698–702 (1999).
8. Borzillo, G. V., Ashmun, R. A. & Sherr, C. J. *Mol. Cell. Biol.* **10**, 2703–2714 (1990).
9. Kondo, M. *et al.* *Nature* **407**, 383–386 (2000).
10. Whetton, A. D. *et al.* *J. Cell Biol.* **125**, 651–659 (1994).
11. Metcalf, D. & Burgess, A. W. *J. Cell. Physiol.* **111**, 275–283 (1982).
12. Dahl, R. *et al.* *Nature Immunol.* **4**, 1029–1036 (2003).
13. Swiers, G., Patient, R. & Loose, M. *Dev. Biol.* **294**, 525–540 (2006).
14. Scadden, D. T. *Nature* **441**, 1075–1079 (2006).
15. Chickarmann, V., Enver, T. & Peterson, C. *PLoS Comput. Biol.* **5**, e1000268 (2009).



Cover illustration
Artwork by N. Spencer

Editor, Nature
Philip Campbell

Publishing
Nick Campbell
Claudia Deasy

Insights Editor
Lesley Anson

Production Editor
Davina Dudley-Moore

Senior Art Editor
Martin Harrison

Art Editor
Nik Spencer

Sponsorship
Amélie Pequignot
Reya Silao

Production
Jocelyn Hilton

Marketing
Elena Woodstock
Emily Elkins

Editorial Assistant
Emma Gibson

TRANSCRIBING THE GENOME

Recent technological advances are transforming our understanding of how the DNA sequence of the genome is transcribed into its functional output of RNA and protein. Researchers are uncovering new layers of complexity on many levels, ranging from the mechanism by which genes are transcribed into RNA to how genetic changes can give rise to disease.

Advances in the detailed mapping of transcription factors across the genome are revealing, for example, unexpected rate-limiting steps during the initiation of gene transcription by RNA polymerase. And such analyses are defining the regulatory regions of genes, as well as the factors that bind to these regions. The distribution of nucleosomes along DNA can now also be finely mapped, showing the dynamic interplay between the packaging of DNA into chromatin and the binding of transcription factors to regulatory regions.

Improved techniques for following the three-dimensional interactions of chromosomes in the nucleus are allowing the effect of such interactions on gene activation and silencing to be explored and are offering glimpses of the poorly understood substructure of the nucleus. Furthermore, high-throughput sequencing methods are uncovering different classes of RNA that are transcribed from various regions of the genome but not translated into proteins, raising the question of the functional importance of these RNAs.

Complementing such experimental methods, computational approaches are revealing the importance of mutations that promote susceptibility to disease not by directly affecting protein-coding sequences but instead by disrupting gene regulatory regions. And computational network approaches are aiding our understanding of the systems-level changes driven by disease-associated mutations.

These Reviews are intended to convey some of the current excitement in the transcription and genomics fields, and we are grateful to the authors and reviewers for their contributions.

Alex Eccleston and Magdalena Skipper, Senior Editors

REVIEWS

186 Defining mechanisms that regulate RNA polymerase II transcription *in vivo*

N. J. Fuda, M. Behfar Ardehali & J. T. Lis

193 The logic of chromatin architecture and remodelling at promoters

B. R. Cairns

199 Genomic views of distant-acting enhancers

A. Visel, E. M. Rubin & L. A. Pennacchio

206 Implications of chimaeric non-co-linear transcripts

T. R. Gingeras

212 Chromosome crosstalk in three dimensions

A. Göndör & R. Ohlsson

218 Molecular networks as sensors and drivers of common human diseases

E. E. Schadt

nature
insight

Defining mechanisms that regulate RNA polymerase II transcription *in vivo*

Nicholas J. Fuda¹, M. Behfar Ardehali¹ & John T. Lis¹

In the eukaryotic genome, the thousands of genes that encode messenger RNA are transcribed by a molecular machine called RNA polymerase II. Analysing the distribution and status of RNA polymerase II across a genome has provided crucial insights into the long-standing mysteries of transcription and its regulation. These studies identify points in the transcription cycle where RNA polymerase II accumulates after encountering a rate-limiting step. When coupled with genome-wide mapping of transcription factors, these approaches identify key regulatory steps and factors and, importantly, provide an understanding of the mechanistic generalities, as well as the rich diversities, of gene regulation.

The genetic information encoded in the DNA of eukaryotic genes is transcribed into RNA by large molecular machines called RNA polymerases. One of these machines, RNA polymerase II (Pol II), transcribes all the protein-coding genes. The control of Pol II activity is highly modulated at individual genes, and this specific regulation is critical for both the homeostasis of cells and the programmed development of multicellular organisms. The execution of this regulation is dictated by combinatorial molecular interactions of transcription factors with each other and with specific DNA sequences at each gene. Modern biochemical and molecular methods coupled with genetics and genomics have identified thousands of factors that participate in regulated transcription¹. Most of these factors are proteins, but a growing number of them are RNAs. They enable Pol II to gain access to the gene's promoter, to initiate RNA synthesis at the transcription start site (TSS) of the gene and to generate a productively elongating transcription complex that produces a full-length RNA transcript.

The thousands of transcription factors involved in the transcription process may be true regulatory factors or simply critical cogs in the cycle of transcription. True regulatory factors are likely to represent only a fraction of the total number of factors that are important for gene expression. As an analogy, consider a motor vehicle: a car has numerous crucial components and processes that are required to achieve acceleration and proper speed (cylinders, spark plugs, tyres and so on), but components regulated by the driver are limited to the ignition, the steering wheel, the accelerator and brake pedals, and the gear stick. Therefore, it is important to identify the true regulatory factors and the associated biochemical processes that execute gene regulation. The status and local density of the ultimate target of regulation, the transcription machine Pol II, have proved extremely useful in assessing the steps in the transcription cycle that are rate limiting and are altered *in vivo* by particular transcription factors (the driver in the above analogy).

In this Review, we discuss how in-depth mechanistic analysis of individual genes coupled with large-scale analysis of transcription-factor binding over an entire genome can distinguish the key steps at which transcription is regulated, and how these steps can be accelerated in an activator-dependent manner.

Gene promoters and factor interactions

The DNA sequences in and around specific gene promoters provide the code that dictates when, where and at what level specific genes are

transcribed. This code comes in three parts: the core promoter, the region proximal to the core promoter, and the more distant enhancer sequences (Fig. 1). In various combinations, the elements of the core promoter sequence target the assembly of distinct preinitiation complexes (PICs) composed of the general transcription factors (GTFs)². Promoter-proximal regions and more distant enhancer sequences direct the binding of specific transcription factors, called activators or repressors (see page 199 for a more detailed discussion of enhancers). Although activators or repressors can interact directly with components associated with the core promoter, they execute their regulation predominantly through co-regulators, which are often multiprotein complexes. Some of the co-regulators can interact directly with Pol II and GTFs and influence expression. Others can reorganize nucleosomes or covalently modify chromatin, and change the chromatin architecture of the gene. This can in turn influence transcription-factor associations and the transcriptional status of Pol II.

Although present evidence suggests that many steps in the transcription process may be rate limiting, the question remains whether these rate-limiting steps are actual points of regulation. To meet this criterion, these steps should be regulated by factors in response to particular physiological, environmental or developmental signals. Although transcription regulatory factors that act as repressors can also modulate specific steps, we focus here on activators, as they seem to predominate as critical modulators of gene expression in eukaryotes.

In-depth analyses of individual genes, or sets of co-regulated genes, have revealed critical mechanistic insights into transcription factors that take part in regulation in response to specific cellular signals. This information, when coupled with more recent large-scale analyses of the associations of such factors over the entire genome (which have been carried out by individual laboratories, as well as by the Encyclopedia of DNA Elements (ENCODE) and modENCODE consortia), allows the generality of particular regulatory mechanisms to be assessed. These genome-wide efforts efficiently appraise the collections of genes that associate with particular transcription factors and thereby define many potential participants in any regulatory mechanism. They also reveal the regulatory circuitry of gene expression networks and how these networks respond to cellular signalling³. Interpreting how the transcription factors and gene circuitry respond to signals and lead to transcriptional regulation requires that we identify not only the factors that respond to signals but also the rate-limiting steps in transcription.

¹Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA.

Rate-limiting steps in transcription

The transcription cycle consists of at least eight distinct major steps at which transcription could be rate limiting and activators could potentially act to increase the rate of transcription (Fig. 2). The transcription cycle begins with Pol II gaining access to the promoter, which in some cases requires the promoter being cleared of nucleosomes that obscure access to Pol II and the GTFs (step 1). A PIC assembles on the core promoter (step 2). The DNA is then unwound, and Pol II initiates transcription (step 3). Early-elongating Pol II gets a stable grip on both the DNA and the growing RNA chain, escapes/clears the core promoter and proceeds to the promoter-proximal pause region (step 4). The paused Pol II complex is then hyperphosphorylated and escapes from the pause region in an unknown manner, either terminating or entering productive elongation (step 5). If it has not terminated, Pol II must then productively elongate through the entire body of the gene (step 6). After this, Pol II undergoes termination (step 7), and it can reinitiate to start a new round of transcription (step 8).

Any of these major steps could, in principle, be rate limiting, and the distribution of Pol II across a gene can suggest which steps are rate limiting for that gene. The Pol II density across many genes has been determined in a plethora of individual gene studies⁴; moreover, a wealth of data has been obtained in recent genome-wide chromatin immunoprecipitation (ChIP) studies examining Pol II distribution across the genomes of several organisms: *Saccharomyces cerevisiae*⁵, *Drosophila melanogaster*^{6,7} and *Homo sapiens*⁸. In each organism, these studies have identified different classes of gene on the basis of their Pol II distribution: no Pol II, Pol II evenly distributed and Pol II enrichment at the 5' end. Genes without Pol II are in an 'off' state, and are limited by the step at which the promoter is cleared of nucleosomes (step 1) or the step at which a PIC assembles (step 2). An even distribution of Pol II suggests that Pol II recruitment (step 2) is the rate-limiting step: none of the downstream steps leads to an accumulation of Pol II in other regions of the gene⁹. An enrichment in Pol II at the 5' end suggests that steps downstream of Pol II recruitment (steps 3–5) are rate limiting. Because ChIP localization with a single Pol-II-specific antibody cannot distinguish between steps 3, 4 and 5, more experiments pinpointing the exact rate-limiting step need to be performed. The transition between PIC formation (step 2) and promoter escape (step 4) is marked by the unwinding of DNA, formation of a transcription bubble with a stable RNA–DNA duplex and lengthening of the nascent transcripts associated with Pol II. Transcription-bubble formation and RNA length can be distinguished by permanganate mapping of unpaired thymidines in the transcription bubble¹⁰ and run-on assays^{11,12}, respectively. In addition, the transition between initiation and pausing (step 4) is marked by

phosphorylation of the Pol II carboxy-terminal domain (CTD) repeats on Ser 5 by the kinase subunit of the GTF TFIIF (CDK7 in *Drosophila*), and productive elongation (step 6) is generally marked by phosphorylation of Pol II CTD repeats on Ser 2 by the kinase complex positive transcription elongation factor b (P-TEFb; CDK9–cyclin T in *Drosophila*). Therefore, using specific antibodies to examine these phosphorylation marks on genes with 5'-end Pol II peaks can help distinguish the rate-limiting step for those genes¹³.

Regulating Pol II recruitment

Many genes regulated by the recruitment of Pol II have promoters covered with nucleosomes. Activators at these genes recruit nucleosome remodellers and nucleosome-modifying enzymes to allow GTFs and Pol II access to the promoter (Fig. 2, step 1) (see page 193 for details on nucleosome remodellers). *PHO5* in *S. cerevisiae* is one of the best studied of the genes that are regulated in this manner (Box 1). In other examples, it has been shown that both human and yeast activators interact with the SWI/SNF remodelling complexes (Swi/Snf complex in yeast) and positively stimulate transcription from nucleosome-containing templates¹⁴. In addition, recruitment of histone-modifying enzymes (for example recruitment of the histone acetyltransferase Gcn5 to galactose-inducible genes by the yeast activator Gal4 (ref. 15)) provides another means by which activators influence and modulate the outcome of transcription by modifying promoter chromatin state.

In other genes, the promoter is free from nucleosomes, but Pol II recruitment is still rate limiting (step 2). During activated transcription, recruited Pol II quickly progresses into productive elongation and becomes relatively uniformly distributed across the gene¹⁶. At these genes, PIC assembly must be upregulated by activators. Extensive *in vitro* studies have shown activators can interact with many GTFs: TATA-binding protein (TBP), TFIID, TFIIA and TFIIB¹⁷. Activators also recruit the coactivator Mediator, which can interact with GTFs and increase expression^{18,19}. These interactions might increase the binding of GTFs to the promoter or stabilize the PIC, allowing more efficient recruitment of Pol II. Additionally, activator-dependent recruitment of chromatin-modifying enzymes results in distinctive chromatin marks on promoters. Domains associated with GTFs can bind to these marks^{20,21}, and these interactions can further aid in stabilizing PIC formation.

Regulating post-recruitment steps

In vivo Pol II distributions have also indicated that post-recruitment steps can be rate limiting. Enrichment in Pol II at the 5' ends of genes suggests that steps between recruitment and productive elongation (steps 3–5) are

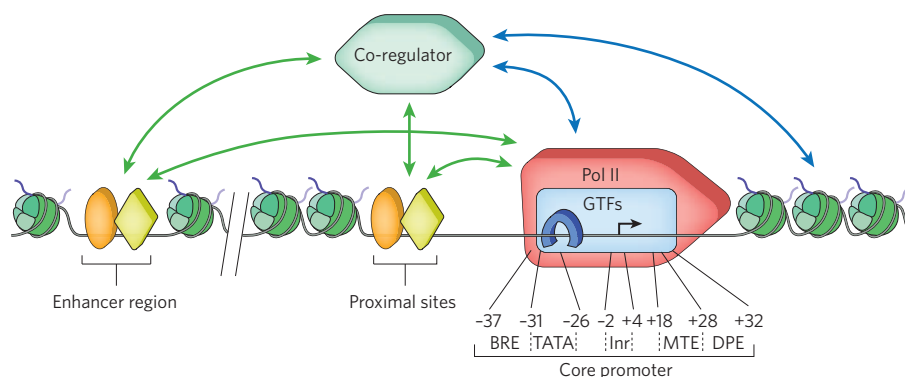


Figure 1 | Transcription regulatory interactions. General transcription factors (GTFs) bind to specific sequence elements in the promoter. These elements (the B recognition element (BRE), the TATA box (TATA), the initiator (Inr), the motif ten element (MTE) and the downstream promoter element (DPE)) and their approximate locations relative to the transcription start site (TSS, black arrow) are shown². Transcriptional regulators (orange oval and yellow diamond), which are either activators or repressors, bind to specific DNA sequences located near the core promoter of the gene or various distant regions, called enhancers. The regulators can interact (green arrows) with

GTFs, such as TFIID (blue rectangle) and TATA-binding protein (TBP, blue horseshoe), and the Pol II complex (red 'rocket') to enhance or repress transcription. They also interact (green arrows) with co-regulators (green hexagon) that can interact (blue arrows) with the general transcription machinery or chromatin-modifying factors, such as histone modifiers or nucleosome remodellers. The co-regulators can also bind to nucleosomes (green) with various histone modifications, stabilizing the co-regulator binding to the gene. Activators can recruit, stabilize or stimulate these factors, and repressors can disrupt or inhibit these factors.

rate limiting in these genes. Some of them may be regulated at initiation (step 3) or promoter escape (step 4). In the case of regulation at initiation, the Pol II associated with the 5' end of the gene is contained within a PIC, and activators may regulate open-complex formation by recruiting or stimulating factors important for this step. For example, Mediator can interact with two GTFs crucial for unwinding DNA and forming open complexes: TFIIE and TFIIH^{18,22}. Therefore, activators recruiting Mediator may increase the rate of open-complex formation. In the case of regulation at promoter escape/clearance, the Pol II associated with the 5' end of the gene has initiated transcription but cannot transcribe to the promoter-proximal pause region owing to the instability of the RNA–DNA duplex and the inability of Pol II to break contacts with factors establishing the PIC. This can lead to abortive initiation⁴. Activators may mitigate these problems, but results on the extent of regulation at step 4 or how this happens *in vivo* are limited so far. TFIIH is again important for this step, not only for further unwinding of downstream DNA but also for the TFIIH-dependent Ser 5 CTD phosphorylation that occurs around this step, which may aid in breaking Pol II contacts with some promoter-bound factors²³. Indeed, an activator can promote this phosphorylation *in vitro*²⁴, and Mediator enhances the TFIIH-dependent phosphorylation of the CTD¹⁹.

Assays other than ChIP have shown that the Pol II that is enriched on the 5' ends of many genes is already engaged in transcription but is held paused¹². Directed studies of specific genes in the 1980s showed that Pol II was at high density on the 5' ends of some genes, and this Pol II was extensively characterized in focused studies of *Drosophila Hsp70*

and other heat-shock genes (Box 2; reviewed in ref. 25). Upon activation, the paused Pol II on *Hsp70* is released into productive elongation, and Pol II becomes evenly distributed across the gene. This indicates that the activator is regulating the transition from the paused state to productive elongation (step 5). P-TEFb is a major switch that has a critical role in facilitating the transition of Pol II from promoter-proximal pause sites into productive elongation²⁶ at most (if not all) genes; inhibition of P-TEFb dramatically decreases global transcription²⁷. P-TEFb interacts directly with some activators^{28–30}, but others rely on different mechanisms to recruit P-TEFb indirectly (reviewed in ref. 31). Although P-TEFb is important for pause escape, Pol II still elongates many dozens of base pairs from the canonical *Hsp70* pause sites when P-TEFb is inhibited during heat shock²⁶. Therefore, there may be other P-TEFb-independent mechanisms for releasing paused Pol II. In addition, elongation requires nucleosome loss or remodelling to occur, and it has been proposed that nucleosomes block the escape from pausing³².

At present, the case for regulation at later stages (steps 6–8) in the transcription cycle is hard to make, but hints of such regulation exist³³. It seems probable, for some genes, that cells have evolved means of at least modest regulation at these stages in response to cellular signals. Activator-dependent loss of nucleosomes aids in elongation (step 6). Additionally, the activator-dependent GTF-stabilizing interactions discussed earlier are important for recycling and reinitiation of Pol II (step 8). Some GTFs can remain associated with the promoter after the Pol II has escaped, and they form a scaffold that allows Pol II to initiate efficiently in successive rounds of transcription³⁴.

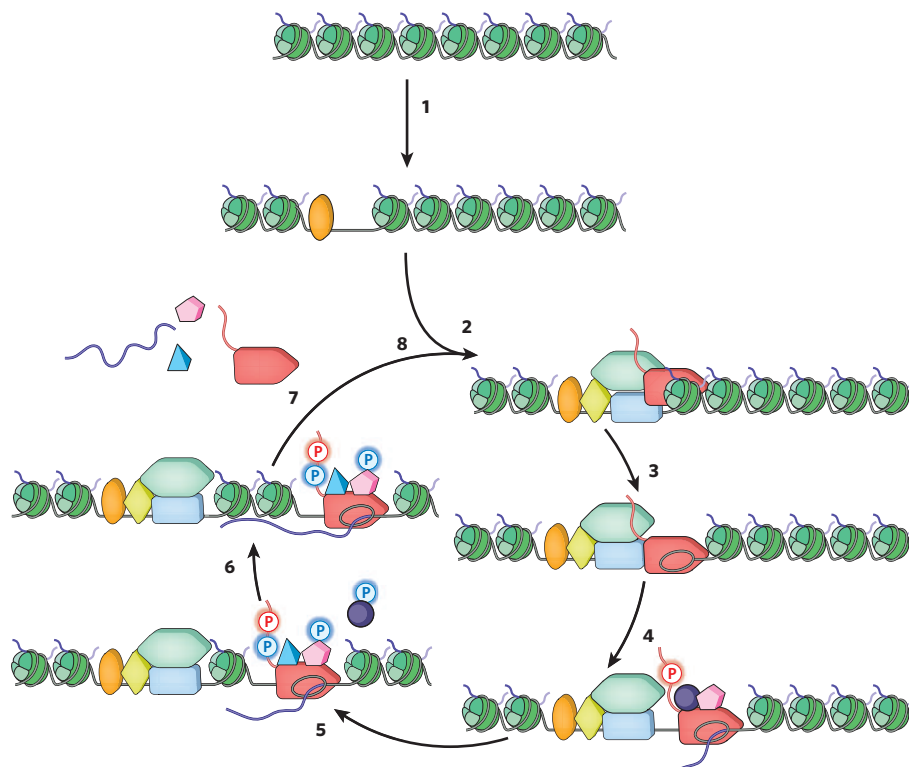


Figure 2 | The transcription cycle is a multistep process. Step 1: chromatin opening. The repressed gene and regulatory region are entirely packaged as nucleosomes (green). An activator (orange oval) binds and recruits nucleosome remodellers to clear the promoter. Step 2: PIC formation. A second activator (yellow diamond) binds, promotes the binding of GTFs (blue rectangle) and recruits coactivators (green hexagon), facilitating Pol II (red rocket) entry to the PIC. Step 3: initiation. DNA is unwound (oval inside Pol II) at the TSS, and an open complex is formed. Step 4: promoter escape/clearance. Pol II breaks contacts with promoter-bound factors, transcribes 20–50 bases downstream of the TSS, produces an RNA (purple line) and pauses, partially mediated by SPT4–SPT5 in *Drosophila* (pink pentagon) and

negative elongation factor (NELF) complex (purple circle). The Ser residues at position 5 (Ser 5) of the Pol II carboxy-terminal domain (CTD) repeats are phosphorylated (red P) during this step. Step 5: escape from pausing. P-TEFb (blue triangle) is recruited directly or indirectly by the activator and phosphorylates Ser 2 of the Pol II CTD repeats, SPT5 and the NELF subunits (blue Ps). NELF dissociates from the rest of the complex. Pol II escapes from the pause, either terminating or entering productive elongation. Step 6: productive elongation. Nucleosomes are disassembled and reassembled as the Pol II elongation complex transcribes through the gene. Step 7: termination. After the Pol II complex transcribes the gene, it is removed from the DNA, and the RNA is released. Step 8: recycling. The freed Pol II can reinitiate.

Box 1 | The *Saccharomyces cerevisiae* *PHO5* gene is regulated at the chromatin-opening step

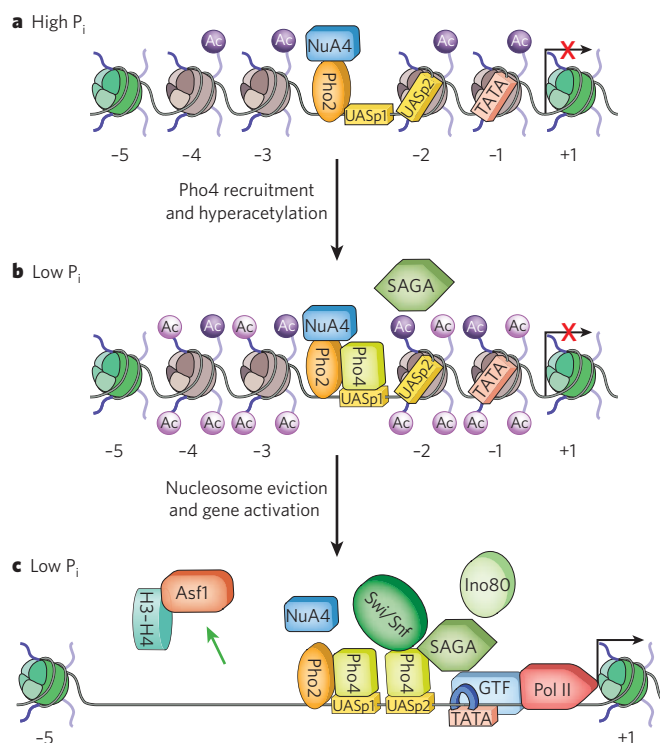
Transcription from the *Saccharomyces cerevisiae* acid-phosphatase gene *PHO5* (see figure, panel **a**) is regulated at the level of activator recruitment and eviction of four positioned nucleosomes (brown, -1 to -4) from the upstream regulatory and promoter region^{51,52}. Pho2, a homeodomain-containing activator, and the histone acetyltransferase complex NuA4, which acetylates histones H4 and H2A (purple Ac) before induction, are both present at the promoter. Phosphate (P_i) starvation (see figure, panel **b**) induces *PHO5* by activating the cyclin-dependent-kinase inhibitor Pho81 (not shown), which inhibits the Pho80–Pho85 kinase complex (also not shown) and allows accumulation of the active unphosphorylated form of the basic helix–loop–helix activator Pho4 in the nucleus⁵³. Pho4 binds mainly to the low-affinity UASp1 within the hypersensitive site that is flanked by two positioned nucleosomes on each side, and cooperatively interacts with Pho2. This Pho4–Pho2 complex triggers disruption of the positioned nucleosomes, and this event is concurrent with Pho4 binding to the high-affinity UASp2 and induction of transcriptional activation in a manner that depends on the acidic transactivation domain of Pho4 and on NuA4 (refs 54–56).

After Pho4 binding, the positioned nucleosomes become hyperacetylated (light mauve Ac) through the histone-acetyltransferase activity of the SAGA subunit Gcn5 and then undergo remodelling (see figure, panel **b**) before being evicted (see figure, panel **c**, green arrow) from the promoter. Both Swi/Snf and Ino80 complexes have been implicated in chromatin remodelling at *PHO5* (refs 56, 57). The H3–H4 histone chaperone Asf1 has also been shown to play a part in the eviction process^{58,59}. Although Gcn5, Asf1 and chromatin remodellers are not essential for *PHO5* induction,

their deletion results in a kinetic delay in the loss of nucleosomes and gene activation. These observations indicate that multiple mechanisms are in place for remodelling and eviction of the positioned nucleosomes at *PHO5*. The co-regulated *PHO8* gene is dependent on Gcn5 and Swi/Snf⁶⁰, indicating that these nucleosome modifications and remodelling events can have a range of effects on the Pho4-mediated activation of this co-regulated gene family.

Other phosphate-responsive genes are also induced during P_i starvation. But the degree

of sensitivity to environmental P_i and the extent of expression on induction vary greatly among these genes. A recent study showed that variabilities in the activation threshold and transcription range of phosphate-system genes are governed, respectively, by the accessibility of high-affinity Pho4-binding sites before induction and the affinity and number of these Pho4-binding sites⁶¹, highlighting the role of activator binding-site accessibility and nucleosome positioning on the dynamic range of transcriptional output.

**Benefits of regulating at different steps**

As suggested from this discussion, activators can act during distinct steps in transcription *in vivo*. Certain activators, such as Sp1 in mammals, target early steps in the cycle, whereas others, such as those with an acidic activation domain, can target early elongation/escape from pausing. Studies suggest that the distinct sets of targets may be independent of one another. The very strong viral acidic activator VP16 seems to act at both early and pausing escape steps³⁵. The effect of *Drosophila* activator HSF on nucleosome removal could be separated from its effects in stimulating transcription on the *Hsp70* gene³⁶. The ability of activators to stimulate multiple slow steps can lead to a much more rapid and robust activation through a kinetic synergism (reviewed in ref. 37).

The different steps in transcription provide multiple targets for the evolution of regulatory mechanisms. A block at early stages of promoter accessibility provides a means of placing a gene under tight control. An activator that stimulates nucleosome removal to unmask the promoter would allow that first step to occur; however, the gene could then require additional activators to stimulate later steps that eventually produce a messenger RNA. Thus, the activation of a gene could be regulated by a combination of signals that each acts on particular activators and their targeted steps, resulting in tight control; an example of such a gene is *PHO5* (Box 1).

The promoter-proximal paused Pol II seems to provide a means of achieving a rapid, and perhaps synchronous, activation of gene expression³⁸. The paused Pol II has already progressed through multiple processes

that can be slow and stochastic, and a transcriptional activator, acting on a preloaded paused Pol II, allows a rapid transition into productive elongation. Genes with paused Pol II seem not to be in a completely transcriptionally 'off' state¹². Therefore, regulation of pausing may sacrifice tight control of RNA production in favour of the uniform and rapid response of a gene. The heat-shock genes are a classic example of this regulation: their rapid induction seems critical in responding to a stress that is normally lethal (Box 2). Other stress-response genes, such as those responsible for DNA-damage, unfolded-protein and immune-response pathways, are also enriched in paused Pol II^{6,12}. In the early embryo, narrow bands of cells must respond rapidly and uniformly to developmental signals, and genes that respond to these signals are also highly enriched in paused Pol II at the developmental stage at which they must be turned on⁷.

A wish list for future approaches

Although many powerful methodologies have been developed for investigating mechanisms of gene regulation *in vivo*, there follows a wish list of key tools and approaches for the future. This list is not meant to be comprehensive, and the approaches described benefit both from the interplay with *in vitro* studies, which provide critical tests of mechanisms and quantification of binding and rate constants for factor interactions, and from structural studies, which provide insight into the precise molecular architectures of proteins and larger macromolecular complexes.

First in the list are protein–DNA crosslinking technologies (for molecular imaging), which produce snapshots of transcription-factor

interactions on specific genes *in vivo* and thereby set critical limits in evaluating models of transcriptional regulation. This approach is especially powerful when applied at high temporal and spatial resolution to track the recruitment kinetics and location of specific factors on specific genes during a time course of their activation^{13,39}. These crosslinking methods can be used at individual genes, as well as at the whole genome level. As the resolution of these assays improves, so too does the power of this data in evaluating mechanistic models of transcriptional regulation. Ultimately, the utility of these assays would benefit from the development of this technology to allow mapping of contacts at single-nucleotide resolution and at sufficient kinetic resolution to resolve known major steps in the transcription cycle (Fig. 2), as well as steps yet to be discovered.

Second are highly sensitive microscopy methods, which should provide a strong complement to biochemical methods for examining protein–DNA interactions by allowing observation of the recruitment and dynamics of proteins in real time. The tracking of factors during a time course of the rapid and synchronous activation of a regulated gene will be greatly enhanced when microscopic imaging technology is sufficiently sensitive to examine the recruitment and dynamics of individual proteins on a single chromatid *in vivo*. Tracking proteins at specific loci is now possible on polytene chromosomes or in diploid

cells, where genes are amplified in tandem, but single-chromatid tracking of Pol II and particular transcription factors would offer a comprehensive and ordered view of the process and provide the detail that is often masked in measurements that rely on averaging events at many gene copies in a single cell or biochemical measurements of genes in a population of cells.

Third are methodologies that evaluate the catalytic and modification state of the key proteins, which, along with the tracking of protein–DNA interactions *in vivo*, are also critical. The antibodies that detect the phosphorylation status of Pol II have been crucial in assessing the activity state of Pol II at various positions along a gene and during the time course of gene activation. Additional antibodies, or other detection reagents, that can evaluate the modification status of transcription factors could certainly provide valuable insights into the way in which different modifications influence each other, and how the final modification code influences the mechanisms of activation. Ultimately, the development of highly effective chromatin purification schemes and highly sensitive mass spectrometry should allow the examination of the complete range of proteins and protein modifications in a particular region and under any condition. There has already been some success in such an examination of a repetitive region of the genome⁴⁰, and taking this to the level of specific genes would be extremely powerful.

Box 2 | The *Drosophila Hsp70* gene is regulated at the pause-escape step

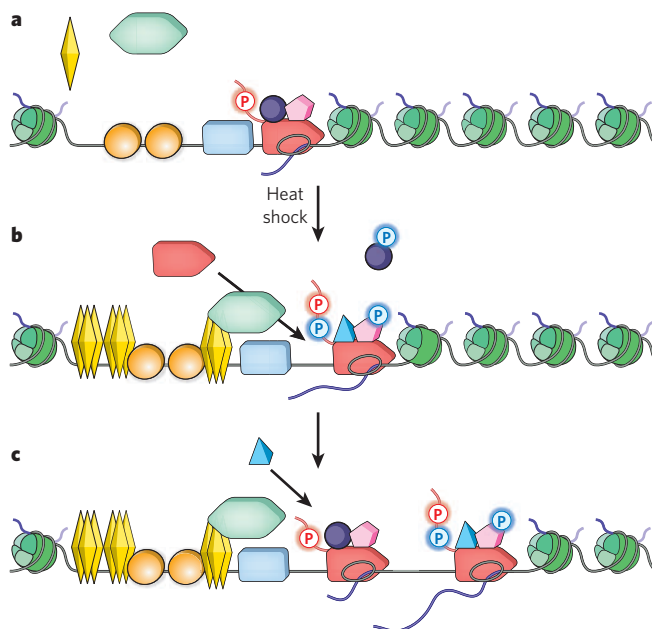
Drosophila Hsp70 was one of the first genes discovered to have promoter-proximal paused Pol II, and has been extensively studied. As a result, *Hsp70* has served as the model for genes regulated at the step of early elongation. Its promoter resides in a nucleosome-free region extending to about 250 bases downstream of the TSS^{36,62} (see figure, panel a). This open promoter is bound by GAGA factor (GAF, orange circles) and GTFs (blue rectangle)⁶³. Studies have suggested that the GAGA elements are crucial for setting up the paused Pol II (red rocket)^{64–66}, which is partially phosphorylated (red P). And *in vitro* evidence suggests that GAF bound to the promoter can recruit nucleosome remodellers to maintain this nucleosome-free state⁶⁷. This open promoter allows Pol II to initiate and transcribe 20–40 bases downstream of the TSS, where it is held paused. This pausing is, at least partially, mediated by the SPT4–SPT5 complex (pink pentagon) and the NELF complex (purple circle). *In vivo*, NELF is present on uninduced *Hsp70*, and it is still present, but at lower levels, after heat shock⁶⁸ (see figure, panel c). Furthermore, NELF depletion *in vivo* reduces the amount of engaged Pol II on uninduced *Hsp70* (ref. 69). Additionally, the downstream sequence may also be important for pausing. When the sequence within 30 bases downstream of the *Hsp70* TSS is switched with the sequence from another gene, the amount of pausing markedly decreases⁶⁴. This may indicate that either the factors binding to downstream elements or the intrinsic pause-inducing characteristics of the transcribed sequence, or both, have a role in pausing. The paused Pol II is phosphorylated by the TFIIH subunit CDK7 on Ser 5 of the CTD repeats. This phosphorylation may be involved in pausing. A temperature-sensitive mutant of CDK7 decreases the amount of paused polymerase on *Hsp70* at non-permissive temperatures⁷⁰; whether this affects pausing

directly or at an earlier step remains to be resolved.

Heat shock (see figure, panel b) causes the transcriptional activator HSF (yellow diamonds) to trimerize and stably bind upstream of *Hsp70* (ref. 71). Such a temperature shift also activates HSF, resulting in the recruitment of coactivators (green hexagon), a rapid general loss of nucleosome protection across the gene³⁶ and release of the paused Pol II into productive elongation. Upon heat shock, P-TEFb (blue triangle) is recruited to the gene^{13,72} and phosphorylates (blue P) the CTD, SPT5 and NELF subunits; the NELF complex dissociates from the Pol II complex; and Pol II releases from the pause sites, allowing rapid recruitment of new Pol II to the gene (see figure, panel c). Although Pol II still resides in the canonical pause sites

under these conditions, it is estimated that the pause is of much shorter duration, with Pol II escaping every 4 s rather than once every 10 min before heat-shock induction⁶³.

Several studies have demonstrated that P-TEFb is important for releasing the paused polymerase upon induction of *Hsp70*. *In vitro* assays show that P-TEFb relieves the inhibitory effects of SPT4–SPT5 and NELF⁷³. Depletion or inhibition of P-TEFb severely reduces *Hsp70* RNA expression^{70,74}, and P-TEFb inhibition, either before or after heat shock, blocks Pol II escape from the 5' end of the gene²⁶. Additionally, TFIIIS is important for Pol II escape from the pause sites through its maintenance of paused Pol II in an elongation-competent state⁷⁵. Depletion of TFIIIS impedes the release of Pol II from the pause and reduces the rate of *Hsp70* mRNA production.



The evaluation of Pol II activity state is enhanced by nuclear run-on assays that measure transcriptionally engaged RNA polymerase complexes. RNA polymerases that are in an elongation state or simply associated with DNA can be detected by ChIP assays, whereas only the former are detected by nuclear run-on assays. An approach called GRO-seq, which uses massively parallel sequencing to measure nascent run-on transcripts, has greatly enhanced the sensitivity of nuclear run-on assays and provides a genome-wide analysis of all transcriptionally engaged polymerases¹². Fourth in our wish list is further development of the GRO-seq assay and the continued examination of short RNAs^{41,42} and RNAs associated with chromatin⁴³, which may allow various states of elongating Pol II to be distinguished (for example the promoter-proximal paused, arrested, abortively initiating and productively elongating states) and further enhance our understanding of the transcription cycle. The utility of these GRO-seq and derivative assays will be enhanced by the development of strategies that allow single-nucleotide resolution and thereby enable the location of Pol II to be precisely defined relative to sequence elements and particular transcription factors.

Although the mapping of protein–DNA interactions *in vivo* at specific genes is well developed, the determination of protein–protein interactions (which are equally important) is much less so and is the fifth item in our wish list. High-resolution microscopy methods that provide subwavelength resolution, for example fluorescence resonance energy transfer⁴⁴ and stochastic optical reconstruction microscopy/photo-activated localization microscopy^{45,46}, have the resolution to assess whether proteins are separated by tens of nanometres or less and thus evaluate whether these proteins are close enough to be in contact. Other optical techniques such as fluorescence cross-correlation spectroscopy⁴⁷ make it possible to assess whether pairs of proteins are part of the same complex. The use of biological amplification provided by polytene chromosomes⁴⁷ or tandem polymers of genes⁴⁸ has provided a glimpse of the potential of optical methods in viewing transcription at specific genes. With improvements in fluorescent labels and detection methods, these approaches should in principle be applicable to factors associated with specific genes, allowing factor–factor associations to be tracked in real time. In addition, recent studies in *S. cerevisiae* have used photoreactive amino acids to provide detailed protein–protein contacts during initiation *in vivo*⁴⁹. Extending this analysis to other steps and other organisms will allow *in vivo* protein–protein interactions to be examined during the transcription cycle in unprecedented detail.

Last, pronounced augmentation of the optical and molecular imaging described here can be achieved by depleting or inhibiting specific factor interactions. Re-examining (re-imaging) the consequences of such experimental treatments can provide critical tests of proposed mechanisms. Although depleting factors with RNA interference is convenient and can generally be used to disrupt factors, sorting primary effects from secondary effects is difficult. Drugs that target specific transcription-factor kinases have been particularly useful, especially when effects are examined immediately after cells have been treated^{23,26}. Ultimately, cell-permeable drugs²⁶, or RNA-aptamer-based drugs⁵⁰ synthesized in cells, that target protein–protein interactions during transcription will be extremely useful for assessing the primary effects of such perturbations.

Outlook

The rapid advancement of techniques in biochemistry and microscopy is providing powerful methods to examine the molecular details of biological processes in living cells. These techniques, when coupled with sophisticated approaches to genetically alter and chemically inhibit transcription factors, will provide a new understanding of the transcription cycle, including more detailed knowledge of the known steps in transcription and perhaps identification of new steps. Analysis of individual genes will continue to reveal important mechanistic information about transcription-factor function. Such studies of single genes will be complemented with genome-wide assays to investigate the generality of discoveries and identify specific mechanisms used by individual and co-regulated genes. The next decade will undoubtedly yield exciting insights into the mechanisms of transcription and its regulation. ■

- Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
- Juven-Gershon, T., Hsu, J. Y., Theisen, J. W. & Kadonaga, J. T. The RNA polymerase II core promoter — the gateway to transcription. *Curr. Opin. Cell Biol.* **20**, 253–259 (2008).
- Lee, T. I. *et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799–804 (2002).
- Saunders, A., Core, L. J. & Lis, J. T. Breaking barriers to transcription elongation. *Nature Rev. Mol. Cell Biol.* **7**, 557–567 (2006).
- Venters, B. J. & Pugh, B. F. A canonical promoter organization of the transcription machinery and its regulators in the *Saccharomyces* genome. *Genome Res.* **19**, 360–371 (2009).
This paper examines the genome-wide distribution of Pol II in yeast and suggests that yeast have Pol II enrichment on the 5' ends of many genes.
- Muse, G. W. *et al.* RNA polymerase is poised for activation across the genome. *Nature Genet.* **39**, 1507–1511 (2007).
This paper examines the genome-wide distribution of Pol II in *Drosophila* S2 cells and demonstrates that there is Pol II enrichment on many promoters in a stalled or paused state.
- Zeitlinger, J. *et al.* RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nature Genet.* **39**, 1512–1516 (2007).
This paper examines the genome-wide distribution of Pol II in *Drosophila* embryos and suggests that promoter-proximal stalling/pausing occurs at many developmental genes.
- Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).
This paper examines the promoter architecture in human cells and suggests that there is Pol II enrichment on promoters and that Pol II seems to have initiated transcription at most genes.
- Ptashne, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386**, 569–577 (1997).
This paper reviews the evidence for regulation by Pol II recruitment in bacteria and yeast.
- Mirkovitch, J. & Darnell, J. Jr. Mapping of RNA polymerase on mammalian genes in cells and nuclei. *Mol. Cell Biol.* **3**, 1085–1094 (1992).
- Rougvié, A. E. & Lis, J. T. The RNA polymerase II molecule at the 5' end of the uninduced *hsp70* gene of *D. melanogaster* is transcriptionally engaged. *Cell* **54**, 795–804 (1988).
- Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
This paper examines the transcriptionally engaged polymerase across the genome in human cells, and demonstrates a 5'-end enrichment in engaged Pol II at many genes.
- Boehm, A. K., Saunders, A., Werner, J. & Lis, J. T. Transcription factor and polymerase recruitment, modification, and movement on *hsp70* *in vivo* in the minutes following heat shock. *Mol. Cell Biol.* **23**, 7628–7637 (2003).
- Peterson, C. L. & Workman, J. L. Promoter targeting and chromatin remodeling by the SWI/SNF complex. *Curr. Opin. Genet. Dev.* **10**, 187–192 (2000).
- Larschan, E. & Winston, F. The *S. cerevisiae* SAGA complex functions *in vivo* as a coactivator for transcriptional activation by Gal4. *Genes Dev.* **15**, 1946–1956 (2001).
- Ahn, S. H., Kim, M. & Buratowski, S. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol. Cell* **13**, 67–76 (2004).
- Stargell, L. A. & Struhl, K. Mechanisms of transcriptional activation *in vivo*: two steps forward. *Trends Genet.* **12**, 311–315 (1996).
- Esnault, C. *et al.* Mediator-dependent recruitment of TFIID modules in preinitiation complex. *Mol. Cell* **31**, 337–346 (2008).
- Kim, Y. J., Bjorklund, S., Li, Y., Sayre, M. H. & Kornberg, R. D. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599–608 (1994).
- Vermeulen, M. *et al.* Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**, 58–69 (2007).
- Jacobson, R. H., Ladurner, A. G., King, D. S. & Tjian, R. Structure and function of a human TAFII250 double bromodomain module. *Science* **288**, 1422–1425 (2000).
- Sakurai, H. & Fukasawa, T. Functional connections between mediator components and general transcription factors of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **275**, 37251–37256 (2000).
- Liu, Y. *et al.* Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex. *Mol. Cell Biol.* **24**, 1721–1735 (2004).
- Spiliarakis, C. *et al.* CIITA regulates transcription onset via Ser5-phosphorylation of RNA Pol II. *EMBO J.* **22**, 5125–5136 (2003).
- Core, L. J. & Lis, J. T. Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science* **319**, 1791–1792 (2008).
- Ni, Z. *et al.* P-TEFb is critical for the maturation of RNA polymerase II into productive elongation *in vivo*. *Mol. Cell Biol.* **28**, 1161–1170 (2008).
- Chao, S. H. & Price, D. H. Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription *in vivo*. *J. Biol. Chem.* **276**, 31793–31799 (2001).
- Eberhardy, S. R. & Farnham, P. J. Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the *cad* promoter. *J. Biol. Chem.* **277**, 40156–40162 (2002).
- Wittmann, B. M., Fujinaga, K., Deng, H., Ogba, N. & Montano, M. M. The breast cell growth inhibitor, estrogen down regulated gene 1, modulates a novel functional interaction between estrogen receptor α and transcriptional elongation factor cyclin T1. *Oncogene* **24**, 5576–5588 (2005).
- Nowak, D. E. *et al.* RelA Ser276 phosphorylation is required for activation of a subset of NF- κ B-dependent genes by recruiting cyclin-dependent kinase 9/cyclin T1 complexes. *Mol. Cell Biol.* **28**, 3623–3638 (2008).
- Peterlin, B. M. & Price, D. H. Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* **23**, 297–305 (2006).
- Mavrich, T. N. *et al.* Nucleosome organization in the *Drosophila* genome. *Nature* **453**, 358–362 (2008).
- Moore, M. J. & Proudfoot, N. J. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136**, 688–700 (2009).
- Yudkovsky, N., Ranish, J. A. & Hahn, S. A transcription reinitiation intermediate that is stabilized by activator. *Nature* **408**, 225–229 (2000).

35. Blau, J. *et al.* Three functional classes of transcriptional activation domain. *Mol. Cell. Biol.* **16**, 2044–2055 (1996).
36. Petesch, S. J. & Lis, J. T. Rapid, transcription-independent loss of nucleosomes over a large chromatin domain at *Hsp70* loci. *Cell* **134**, 74–84 (2008).
37. Herschlag, D. & Johnson, F. B. Synergism in transcriptional activation: a kinetic view. *Genes Dev.* **7**, 173–179 (1993).
The paper proposes a role for kinetic synergism in the mechanism of transcriptional activation.
38. Boettiger, A. N. & Levine, M. Synchronous and stochastic patterns of gene activation in the *Drosophila* embryo. *Science* **325**, 471–473 (2009).
39. Mason, P. B. & Struhl, K. Distinction and relationship between elongation rate and processivity of RNA polymerase II *in vivo*. *Mol. Cell* **17**, 831–840 (2005).
40. Dejardin, J. & Kingston, R. E. Purification of proteins associated with specific genomic loci. *Cell* **136**, 175–186 (2009).
41. Seila, A. C. *et al.* Divergent transcription from active promoters. *Science* **322**, 1849–1851 (2008).
42. Preker, R. *et al.* RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**, 1851–1854 (2008).
43. Wu, J. & Schibler, U. Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol. Cell. Biol.* **14**, 7219–7225 (1994).
44. Bhaumik, S. R., Raha, T., Aiello, D. P. & Green, M. R. *In vivo* target of a transcriptional activator revealed by fluorescence resonance energy transfer. *Genes Dev.* **18**, 333–343 (2004).
45. Patterson, G. H. & Lippincott-Schwartz, J. Selective photolabeling of proteins using photoactivatable GFP. *Methods* **32**, 445–450 (2004).
46. Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, 810–813 (2008).
47. Yao, J., Munson, K. M., Webb, W. W. & Lis, J. T. Dynamics of heat shock factor association with native gene loci in living cells. *Nature* **442**, 1050–1053 (2006).
48. Janicki, S. M. *et al.* From silencing to gene expression: real-time analysis in single cells. *Cell* **116**, 683–698 (2004).
49. Chen, H.-T., Warfield, L. & Hahn, S. The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. *Nature Struct. Mol. Biol.* **14**, 696–703 (2007).
50. Shi, H., Hoffman, B. E. & Lis, J. T. RNA aptamers as effective protein antagonists in a multicellular organism. *Proc. Natl Acad. Sci. USA* **96**, 10033–10038 (1999).
51. Boeger, H., Griesenbeck, J., Strattan, J. S. & Kornberg, R. D. Nucleosomes unfold completely at a transcriptionally active promoter. *Mol. Cell* **11**, 1587–1598 (2003).
52. Svaren, J. & Hörz, W. Transcription factors vs nucleosomes: regulation of the *PHO5* promoter in yeast. *Trends Biochem. Sci.* **22**, 93–97 (1997).
53. O'Neill, E. M., Kaffman, A., Jolly, E. R. & O'Shea, E. K. Regulation of *PHO4* nuclear localization by the *PHO80*–*PHO85* cyclin–CDK complex. *Science* **271**, 209–212 (1996).
54. Barbaric, S., Munsterkotter, M., Goding, C. & Horz, W. Cooperative *Pho2*–*Pho4* interactions at the *PHO5* promoter are critical for binding of *Pho4* to *UASp1* and for efficient transactivation by *Pho4* at *UASp2*. *Mol. Cell. Biol.* **18**, 2629–2639 (1998).
55. Svaren, J., Schmitz, J. & Horz, W. The transactivation domain of *Pho4* is required for nucleosome disruption at the *PHO5* promoter. *EMBO J.* **13**, 4856–4862 (1994).
56. Reinke, H. & Hörz, W. Histones are first hyperacetylated and then lose contact with the activated *PHO5* promoter. *Mol. Cell* **11**, 1599–1607 (2003).
57. Steger, D. J., Haswell, E. S., Miller, A. L., Wente, S. R. & O'Shea, E. K. Regulation of chromatin remodeling by inositol polyphosphates. *Science* **299**, 114–116 (2003).
58. Korber, P. *et al.* The histone chaperone Asf1 increases the rate of histone eviction at the yeast *PHO5* and *PHO8* promoters. *J. Biol. Chem.* **281**, 5539–5545 (2006).
59. Adkins, M. W., Howar, S. R. & Tyler, J. K. Chromatin disassembly mediated by the histone chaperone Asf1 is essential for transcriptional activation of the yeast *PHO5* and *PHO8* genes. *Mol. Cell* **14**, 657–666 (2004).
60. Barbaric, S. *et al.* Redundancy of chromatin remodeling pathways for the induction of the yeast *PHO5* promoter *in vivo*. *J. Biol. Chem.* **282**, 27610–27621 (2007).
61. Lam, F. H., Steger, D. J. & O'Shea, E. K. Chromatin decouples promoter threshold from dynamic range. *Nature* **453**, 246–250 (2008).
62. Wu, C. The 5' ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I. *Nature* **286**, 854–860 (1980).
63. Lis, J. Promoter-associated pausing in promoter architecture and postinitiation transcriptional regulation. *Cold Spring Harb. Symp. Quant. Biol.* **63**, 347–356 (1998).
64. Lee, H., Kraus, K. W., Wolfner, M. F. & Lis, J. T. DNA sequence requirements for generating paused polymerase at the start of *hsp70*. *Genes Dev.* **6**, 284–295 (1992).
65. Wang, Y. V., Tang, H. & Gilmour, D. S. Identification *in vivo* of different rate-limiting steps associated with transcriptional activators in the presence and absence of a GAGA element. *Mol. Cell. Biol.* **25**, 3543–3552 (2005).
66. Lee, C. *et al.* NELF and GAGA factor are linked to promoter-proximal pausing at many genes in *Drosophila*. *Mol. Cell. Biol.* **28**, 3290–3300 (2008).
67. Tsukiyama, T., Becker, P. B. & Wu, C. ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature* **367**, 525–532 (1994).
68. Wu, C. H. *et al.* Molecular characterization of *Drosophila* NELF. *Nucleic Acids Res.* **33**, 1269–1279 (2005).
69. Wu, C.-H. *et al.* NELF and DSIF cause promoter proximal pausing on the *hsp70* promoter in *Drosophila*. *Genes Dev.* **17**, 1402–1414 (2003).
70. Schwartz, B. E., Larochelle, S., Suter, B. & Lis, J. T. Cdk7 is required for full activation of *Drosophila* heat shock genes and RNA polymerase II phosphorylation *in vivo*. *Mol. Cell. Biol.* **23**, 6876–6886 (2003).
71. Ardehali, M. B. *et al.* Spt6 enhances the elongation rate of RNA polymerase II *in vivo*. *EMBO J.* **28**, 1067–1077 (2009).
72. Lis, J. T., Mason, P., Peng, J., Price, D. H. & Werner, J. P-TEFb kinase recruitment and function at heat shock loci. *Genes Dev.* **14**, 792–803 (2000).
73. Renner, D. B., Yamaguchi, Y., Wada, T., Handa, H. & Price, D. H. A highly purified RNA polymerase II elongation control system. *J. Biol. Chem.* **276**, 42601–42609 (2001).
74. Ni, Z., Schwartz, B. E., Werner, J., Suarez, J. R. & Lis, J. T. Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. *Mol. Cell* **13**, 55–65 (2004).
75. Adelman, K. *et al.* Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. *Mol. Cell* **17**, 103–112 (2005).

Acknowledgements We thank members of the Lis laboratory for discussions and critical reading of the manuscript. Work in our laboratory is supported by US National Institutes of Health grant GM25232.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence should be addressed to J.T.L. (jtl10@cornell.edu).

The logic of chromatin architecture and remodelling at promoters

Bradley R. Cairns¹

The regulation of gene transcription involves a dynamic balance between packaging regulatory sequences into chromatin and allowing transcriptional regulators access to these sequences. Access is restricted by the nucleosomes, but these can be repositioned or ejected by enzymes known as nucleosome remodellers. In addition, the DNA sequence can impart stiffness or curvature to the DNA, thereby affecting the position of nucleosomes on the DNA, influencing particular promoter 'architectures'. Recent genome-wide studies in yeast suggest that constitutive and regulated genes have architectures that differ in terms of nucleosome position, turnover, remodelling requirements and transcriptional noise.

In eukaryotic cells, DNA is wrapped around histone octamers to form nucleosomes, the primary unit of chromatin structure. Nucleosomes compact the genome but also restrict the access of DNA-binding transcription factors, so there is a balance to strike between effective genome packaging and accessibility¹. But cells do more than meet this challenge — they also tailor the way that the chromatin is packaged to help regulate gene expression. This mode of regulation involves dynamic competition between nucleosomes and transcription factors for important *cis*-regulatory sequences in gene promoters. This competition is influenced by enzymes that covalently modify nucleosomes, termed 'chromatin modifiers', and enzymes that reposition, reconfigure or eject nucleosomes, termed 'chromatin remodellers'^{2–4}. Together, these factors help create promoter architectures — defined here as the density, composition and positioning of nucleosomes relative to important *cis*-regulatory sites. These factors also collaborate to alter promoter architecture to expose regulatory sites and allow activation under the appropriate conditions. Furthermore, recent genome-wide studies of nucleosome occupancy^{5–13} and new computational approaches for predicting nucleosome positioning^{14–17} have reinforced earlier notions that the biophysical properties of promoter DNA also help shape the landscape of nucleosome positioning and density at both repressed and active promoters.

Here, I consider this interplay and describe the basic logic for the regulation and remodelling of chromatin architecture at RNA polymerase II (Pol II) promoters. I review recent genome-wide studies on nucleosome occupancy and dynamics in yeast, which have revealed that a substantial minority of genes conform to two contrasting promoter architectures^{18,19} that drive either constitutive or highly regulated genes. I discuss these contrasting architectures and the different tools used to build and regulate them (Box 1) and make comparisons to promoter structures in higher eukaryotes. The underlying concepts can help us understand the regulation of these promoters, as well as 'blended' promoters, which incorporate features of both contrasting architectures.

Two contrasting promoter architectures

The locations and density of nucleosomes and nucleosome variants have been determined across whole genomes in several organisms and cell types^{5–7,9–11,20–22}. Studies in yeast suggest that some promoters (albeit a minority) can be classified into two contrasting architectural categories^{18,19}, 'open' and 'covered' (defined further later), which drive the

two broad types of gene, constitutive and highly regulated, respectively (Table 1 and Fig. 1). These two architectures are contrasting extremes, and many promoters contain a 'blend' of attributes of the two architectures, as well as more complex strategies, to achieve proper regulation. Even so, a discussion of the properties of these two contrasting promoter types is useful for understanding the logic of promoter architecture, and for understanding how and why other promoters might blend attributes of these archetypes to tune their regulation.

Constitutive genes have 'open' promoters

There is dynamic competition between nucleosomes and transcription factors at many promoters. However, constitutive genes have features that favour the binding of transcription factors at the expense of nucleosomes (Fig. 1), and I term these promoters 'open' promoters. Constitutive genes typically have a large (~150-base-pair (bp)) nucleosome-depleted region (NDR) directly upstream of the transcription start site (TSS), within which key *cis*-regulatory sequences reside. This region has traditionally been termed a 'nucleosome-free' region, but there is actually a gradient of depletion, rather than a total loss, so NDR is a more appropriate term. Importantly, the NDR contains poly(dA:dT) tracts²³, DNA sequences that resist bending and so deter nucleosome formation and stability¹⁶. By contrast, AA/TT dinucleotides repeating every 10 bp (with GC dinucleotides 5 bp out of phase) impose a curvature favourable to nucleosome formation and stability, and when extending across a ~150-bp region can function as a nucleosome positioning sequence (NPS)^{14,17,24}. However, the poly(dA:dT) tract is generally considered a more important driver of translational positioning than the NPS¹⁶. Open promoters often combine these two sequence elements into a tripartite structure: a central poly(dA:dT)-rich tract, which deters nucleosome binding, flanked by two NPS elements, which help fix the positions of the two flanking nucleosomes, termed the –1 and +1 nucleosomes in yeast^{5–7,9,25} (Fig. 1). (For human genes, the –1 nucleosome is the one positioned directly upstream of the TSS.) The juxtaposition of both an excluding and a positioning element may help create an exceptionally sharp nucleosome boundary, which deters the encroachment of nucleosomes into the NDR. Importantly, the positions of the +1 and –1 nucleosomes at open promoters align quite precisely with the locations of nucleosomes predicted computationally by programs entrained with NPSs and/or poly(dA:dT) tracts^{14–17}. Indeed, the latest

¹Howard Hughes Medical Institute, Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah 84112, USA.

Box 1 | Chromatin concepts at promoters

The DNA sequence influences the chromatin landscape. DNA curvature or stiffness influences the positioning and density of nucleosomes. Within that landscape, specific sequences define the binding sites for sequence-specific transcription factors, the TATA box and the transcription start site.

Nucleosomes block the access of transcription factors to DNA.

Binding sites located near the middle of a nucleosome are generally inaccessible to transcription factors; those near the edge are partially accessible; and those in the linker are accessible. Inaccessibility leads to a dependence on chromatin remodelling.

Histone chaperones regulate nucleosome dynamics. Histone chaperones assist in both the deposition and the removal of promoter nucleosomes and are partly specialized to function either during or outside replication.

Chromatin remodellers alter nucleosomes. Specialized remodellers organize nucleosome arrays to promote repressed or basal chromatin states (ISWI family, NURF excepted), or incorporate histone variants to promote activation (SWR1 family), or slide and eject nucleosomes to promote DNA access (SWI/SNF family). Some remodellers are targeted or regulated by histone modifications.

'Pioneer' transcription factors typically have a binding site either in the nucleosome-depleted region or between positioned nucleosomes.

This is needed to allow initial access by a transcription factor, but additional binding sites may be hidden under nucleosomes and require a remodeller for access. Alternatively, some pioneer factors can bind sites on the nucleosome surface.

Transcription factors recruit histone modifiers. Modifiers such as acetyltransferases, methyltransferases and kinases can promote or deter the targeting or activity of chromatin remodellers on the proper nucleosome. They also promote or deter the binding of additional chromatin and transcriptional regulatory proteins.

H2A.Z resides at many promoters and positively regulates

transcription. This histone variant has a unique amino-terminal tail that is acetylated when a gene is active. H2A.Z nucleosomes may also be less stable than canonical nucleosomes, allowing ejection and helping to expose binding sites in the promoter.

The +1 nucleosome may help regulate either initiation or promoter-proximal pausing of Pol II. A common feature at inactive promoters in metazoans is the presence of a paused Pol II at the +1 nucleosome.

computational models use positioning information from the reconstitution of nucleosomes on whole genomic DNA *in vitro* to predict the positions of most nucleosomes *in vivo*¹⁵. This underscores the tremendous advances made recently in computational models for nucleosome positioning. Notably, the depth of the NDR is significantly less in human cells than in yeast¹³, perhaps because humans transcribe a much smaller fraction of their genome every cell cycle than yeast cells.

At open promoters, binding sites for transcriptional activators often reside within the NDR itself, not buried under nucleosomes far upstream²³, and their exposure in the NDR promotes activator binding and gene expression. Notably, open promoters in yeast are commonly linked to essential genes and bound by activators that are themselves essential for viability (Table 1). Two other features are correlated with open promoters: a nucleosome containing the histone H2A variant H2A.Z (termed Htz1 in yeast) at the +1 or -1 nucleosome, and a paucity of TATA sequences (discussed further later), the binding site for the TATA-binding protein (TBP) (Table 1 and Fig. 1).

Regulated genes have 'covered' promoters

At regulated genes in their repressed state, nucleosomes often cover the TSS, the regions flanking the TSS, and most of the binding sites for transcriptional activators. Such promoters are hereafter termed 'covered' promoters — a more meaningful term than 'closed' promoters, as nucleosomes cover the proximal promoter. At covered promoters, nucleosomes compete effectively with transcription factors for occupancy of key *cis*-regulatory binding sites, rendering covered promoters more reliant than open promoters on chromatin remodelling and

modifying enzymes to help 'uncover' *cis* sites and allow activity. Covered promoters typically contain NPS elements of varying strength, which help position nucleosomes over binding sites for transcription factors (Fig. 1). However, at least one binding site is typically exposed in the linker DNA between nucleosomes, or partly exposed at the nucleosome edge²³. This exposed site allows a 'pioneer' transcription factor access to the promoter, but chromatin modification and remodelling are probably required to expose the additional sites under nucleosomes³ — a two-step model for activation (Fig. 1).

A classic example of a covered promoter is that of the yeast *PHO5* gene, which has one exposed site for the activator Pho4 between two nucleosomes, and places other Pho4 sites within nucleosomes^{26,27}. Recent work on *PHO5* promoter derivatives shows the remarkable diversity of response dynamics that can be acquired by subtly varying the location and affinity of binding sites for Pho4 within *PHO5* promoter architecture²⁸. With these altered *PHO5* promoters, the affinity of the exposed Pho4 site determines the initial threshold of transcription factor (Pho4) abundance needed for promoter occupancy and initial activation — only promoters with high-affinity exposed sites are occupied and turned on with intermediate Pho4 levels — whereas the affinity of the nucleosome-occluded site(s) sets the upper end of the dynamic range of activation. By extrapolation, it is easy to imagine how combinations of sites for different activator proteins might help tune promoters to respond to multiple cellular signals. Interestingly, a subset of pioneer factors (such as the glucocorticoid receptor) do not require exposure in the linker and, instead, can bind their cognate sites on the nucleosome surface as they bind to only one face of the DNA, and can accommodate nucleosomal DNA curvature²⁹; this allows promoter binding without having prescribed nucleosome positions.

Covered promoters also differ from open promoters in some features of transcription initiation. The transcription initiation factor TBP is present at all Pol II promoters and is required for initiation at TATA-containing and TATA-less promoters. A TATA box is present at ~20–25% of yeast genes and is more highly enriched at covered promoters than open promoters, and is more enriched at highly regulated genes than constitutive genes^{30,31} (Table 1). In yeast, the distance from the TATA box to the TSS can vary from ~25 bp (observed in vertebrates) to ~125 bp. Notably, the relatively few open promoters that contain a TATA box place it quite close to the TSS — typically within 50 bp and clearly within the NDR. By contrast, TATA boxes at covered promoters are more variable in their placement, and the TATA box typically resides inside the edge of the proximal nucleosome, providing partial blockage (Fig. 1). This variable but covered placement of the TATA box at covered promoters reinforces the requirement for chromatin remodelling for TATA exposure, analogous to the remodelling requirement for transcription-factor binding-site exposure. Remarkably, at the yeast *PHO5* promoter, moving the TATA box only a few bases inside or outside the nucleosome edge greatly changes its reliance on chromatin remodelling³². Notably, TATA-containing genes use the TBP-containing complex TFIID, whereas TATA-less promoters rely on a different TBP-containing complex, termed SAGA in yeast and the pCAF/GCN5 complex in humans^{30,31}. SAGA/pCAF complexes contain multiple factors that interact with basal transcription factors, as well as histone-modifying proteins such as the histone acetyltransferase (HAT) Gcn5/pCAF³³. These activities may modify H2A.Z and other promoter proximal nucleosomes to promote nucleosome movement or ejection, promoting constitutive transcription. Notably, when considering the entire repertoire of chromatin remodelling and modifying factors (particularly in yeast), covered promoters are more reliant on chromatin remodelling factors than are open promoters^{18,19}.

Specialized chromatin remodellers

Remodellers have an important role at promoters, helping to construct the initial chromatin states and catalysing transitions to alternative chromatin states (Fig. 2), using the energy from ATP hydrolysis^{2,4,34,35}. Remodellers are specialized multiprotein machines that can be classified by their main functions (Fig. 2): remodellers that belong to the ISWI

Table 1 | Attributes correlated with open or covered promoters

Promoter type	Poly(dA:dT) tracts	Many TF binding sites	TATA box	H2A.Z variant	Histone turnover	Transcriptional plasticity	Transcriptional noise	Chromatin regulation	Condition regulation	Essential genes
Open	++	–	–	++	–	–	–	–	–	++
Covered	–	++	++	–	++	++	++	+	++	–

Correlations: ++, strongly positive; +, moderately positive; –, moderately negative; – –, strongly negative. TF, transcription factor.

family (except NURF and Isw1b) help conduct chromatin assembly and organization^{36,37}, leading to the consistent spacing of nucleosomes; those in the SWI/SNF family provide access to nucleosomal DNA through nucleosome movement³⁸ or ejection³⁹; and those in the SWR1 family reconstruct nucleosomes⁴⁰ by inserting the histone variant H2A.Z into nucleosomes, thereby specializing the composition of nucleosomes. For brevity, I do not discuss the specialized functions of the CHD and INO80 families of remodellers.

ISWI remodellers organize nucleosomes

Apart from NURF and Isw1b, remodellers in the ISWI family carry out nucleosome organization^{36,37}, which often promotes repression. ISWI complexes generally remodel nucleosomes that lack acetylation (at H4K16; that is, Lys 16 of histone H4)⁴¹, confining their activity to nucleosomes at transcriptionally inactive regions. They space nucleosomes by ‘measuring’ the DNA linker between nucleosomes and sliding the nucleosome until the linker DNA reaches a fixed distance, creating nucleosome arrays of uniform spacing^{2–4,42}. In yeast, Isw2 localizes to gene 3′ ends and adjacent to transfer RNAs, and possibly to the –1 nucleosome as well^{12,43}. Chromatin organization by yeast Isw2 helps prevent antisense transcription from occurring in intergenic regions, as well as preventing ‘cryptic’ initiation by Pol II, which can occur if nucleosome density and positioning are not optimized⁴⁴.

Notably, yeast Isw2 can move nucleosomes onto unfavourable DNA sequence elements⁴⁵, which can help establish repression. The yeast *POT1* promoter is a good example of a blended promoter at which this property is illustrated; it is highly regulated but contains a poly(dA:dT) tract and *cis*-regulatory sites that in the repressed state partly overlap with a nucleosome. The omission of Isw2 causes the nucleosome to move away from its overlap with the poly(dA:dT) sequence to a location favoured by sequence alone; this movement exposes binding sites in the promoter and is correlated with partial derepression of the gene⁴⁵. This suggests that certain promoters may use ISWI remodellers to move nucleosomes to unfavourable DNA positions, which could occlude transcription-factor binding sites or the TATA box to confer repression. How might such a promoter be activated in normal cells in the presence of Isw2? Here, the modification of the H4 tail may prevent the action of Isw2 or attract an alternative remodeller (discussed later) that moves the nucleosome away from the poly(dA:dT) tract, exposing a regulatory *cis* element and allowing activation. This example illustrates how the concepts used to regulate open and covered promoters, along with action from the remodeller, can help us understand the regulation of a blended promoter.

SWI/SNF remodellers disorganize nucleosomes

Remodellers in the SWI/SNF family can both slide and eject nucleosomes, and their functions are often correlated with nucleosome disorganization and promoter activation^{2,4,39}. In keeping with this notion, histone acetylation is correlated with gene activity, and SWI/SNF remodellers have domains that bind acetylated tails, promoting their targeting or activity in promoters undergoing activation² — an area of active study beyond the scope of this Review. SWI/SNF action may be needed at open promoters to help remove nucleosomes from poly(dA:dT) sequences or at covered and blended promoters to slide or eject nucleosomes, an important area of future study. In yeast, remodellers in the SWI/SNF family are usually located at the –1 nucleosome¹², consistent with the observation that the binding sites for many condition-specific activators reside within the –1 nucleosome in regulated yeast genes. However, as these remodellers both bind and eject nucleosomes, they are not detected

at NDRs by current methods, perhaps because they remove their binding substrate, the nucleosome. So it remains possible that these remodellers help generate the NDR and regulate the occupancy and position of the –1 nucleosome. This idea is supported by the analysis of mutations in RSC, a member of the SWI/SNF family in yeast; *rsc* mutations are found to affect nucleosome density and positioning at promoters^{46–48}.

An issue of current debate is whether DNA sequences or chromatin remodelling and modifying factors are the most important drivers of nucleosome depletion at promoters. Genome-wide nucleosome occupancy maps generated through the *in vitro* assembly of nucleosomes with yeast genomic DNA show NDRs that map to promoters, but these NDRs are not as depleted as NDRs *in vivo*¹⁵. This leaves testable roles for remodellers, modifiers and variants in NDR creation. Also, roles for SWI/SNF-family members in gene repression have been noted^{2–4}, and here it will be interesting to determine whether nucleosome disorganization promotes the binding of repressors to *cis*-acting sites in enhancers or promoters. A future challenge for studying both SWI/SNF and ISWI remodellers is to isolate *in vivo* remodelling intermediates to provide direct evidence for the involvement of the remodeller in sliding or ejecting particular nucleosomes.

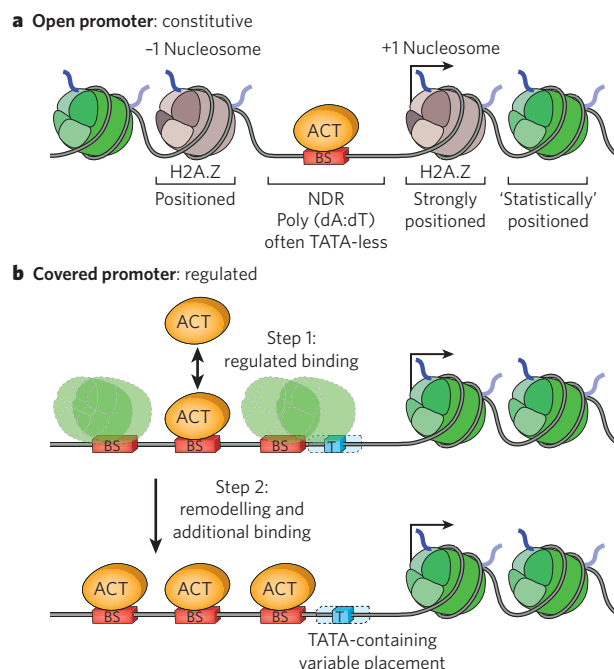


Figure 1 | Properties of open and covered promoters. **a**, Open promoters have a depleted proximal nucleosome adjacent to the transcription start site (TSS, black arrow), a feature common at constitutive genes. **b**, Covered promoters have a nucleosome adjacent to the TSS in their repressed state, a feature common at highly regulated genes. The figure depicts features more common in each contrasting promoter type, but most yeast genes blend the features shown to provide appropriate regulation. Green nucleosomes contain canonical H2A, whereas brown nucleosomes bear H2A.Z. Binding sites (BS) for transcriptional activators (ACT) are shown. These are mainly exposed for open promoters and mainly occluded by nucleosomes (in the repressed state) at covered promoters. Covered promoters typically have nucleosome positioning sequence elements of varying strength and locations that help define nucleosome positions (faded green) and promoter architecture. NDR, nucleosome-depleted region.

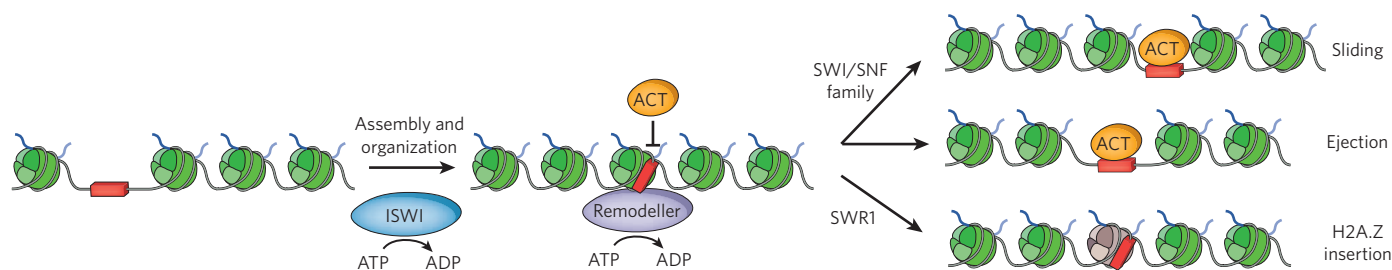


Figure 2 | Basic functions of chromatin remodellers in nucleosome dynamics. Remodellers use ATP hydrolysis to alter nucleosomes and are specialized for certain tasks. Most remodellers of the ISWI family (except NURF and Isw1b) help conduct chromatin assembly and organization and provide consistent spacing of nucleosomes. This organization can cover a binding site (red) for a transcriptional activator (ACT). SWI/SNF-family

remodellers provide access to binding sites in nucleosomal DNA, mainly through nucleosome movement or ejection. SWR1-family remodellers reconstruct nucleosomes by inserting the histone variant H2A.Z into nucleosomes, specializing their composition. This can create an unstable nucleosome in certain compositional and temporal contexts, and might lead to ejection, sliding or reconstruction at promoters.

SWR1 and properties of H2A.Z nucleosomes

The histone H2A variant H2A.Z differs from canonical H2A in its amino-terminal tail sequence and also at key internal residues that may affect its interactions with itself and with the H3/H4 tetramer in the nucleosome, thereby affecting nucleosome stability. H2A.Z is assembled into particular promoter nucleosomes, replacing canonical H2A in a replication-independent manner in a reaction conducted by remodellers of the SWR1 family⁴⁰. There is variation in the placement of H2A.Z within the promoter in different organisms and in its precise role, although certain themes have emerged. In yeast, H2A.Z (Htz1) is found at most genes and mainly occupies the +1 and -1 nucleosome, with lower amounts at +2 (refs 20–22). H2A.Z is highly enriched at open TATA-less promoters²² but is not exclusive to open promoters (many blended promoters contain H2A.Z). In *Drosophila*, H2A.Z is absent at the -1 position but is highly prevalent at the +1 nucleosome, and then likewise decays into the coding region¹¹. In humans, H2A.Z is localized to the promoter, but more widely than in yeast, extending from -3 to +3 in genes with low expression¹³. In plants, H2A.Z is found in the promoter, where it promotes transcription, and is reciprocal with DNA methylation⁴⁹. Functional studies in yeast support roles for H2A.Z in promoting activation^{20–22}, and H2A.Z nucleosomes are lost from genes as transcription increases^{20,22}, a feature also found in humans, where there is profound loss at the -1 position¹³. However, the presence of acetylated H2A.Z increases with activation^{50,51}. This suggests that, in both yeast and humans, inactive or basal genes contain a high level of H2A.Z and that both acetylation and loss of H2A.Z accompanies activation. By contrast, in flies, H2A.Z occupancy at the +1 position correlates with transcription rate¹¹.

At first glance, these contrasting observations suggest very different strategies for H2A.Z. Alternatively, H2A.Z may promote activation by different mechanisms, reflecting differences in the way each organism uses the TSS and the +1 nucleosome in gene regulation. In yeast, the TSS is normally placed just inside the +1 nucleosome, whereas flies tend to have the TSS in the NDR (50–75 bp upstream of the +1 nucleosome)^{11,18}. Furthermore, repressed fly genes often have Pol II already present near the TSS, which may be transcriptionally engaged but paused at the transition to processive elongation (see below and page 186). By contrast, paused Pol II occurs much less frequently in yeast¹². H2A.Z may therefore be poised to regulate transcription initiation in yeast (TSS access), whereas in flies it may help to regulate elongation or pausing at the +1 nucleosome. Another consideration is the timing of H2A.Z placement; yeast genes place H2A.Z into genes in the repressed/basal state, whereas in flies H2A.Z may be inserted after the initial round of transcription to facilitate subsequent rounds¹¹. This may underlie the observation that H2A.Z is lost from yeast genes during activation^{20,22}, whereas H2A.Z abundance increases at the +1 position in flies¹¹. However, in both organisms the strategy is to promote transcription, although by different processes.

Another notable feature of nucleosomes containing H2A.Z is their differential stability, which depends on the histone H3 subtype in the assembled nucleosome. H2A.Z nucleosomes are more stable than

H2A-containing nucleosomes when co-assembled with canonical histone H3 but less stable when co-assembled into nucleosomes with the H3 variant H3.3 (ref. 52). Notably, H3.3-containing nucleosomes are inserted into genes that lose nucleosomes during transcription or chromatin remodeling⁵³ and may therefore contain H2A.Z/H3.3 hybrid nucleosomes at the +1 position after the first round of transcription; this may explain why H2A.Z levels correlate with transcription in flies. Here, the instability (and/or modification status) of the hybrid nucleosome might promote reinitiation or Pol II elongation during subsequent rounds of transcription⁴⁹. By contrast, yeast have only H3.3 (their 'canonical' H3 is the orthologue of vertebrate H3.3), making all yeast H2A.Z nucleosomes slightly unstable, perhaps underlying their depletion during activation⁴⁹. Furthermore, H2A.Z instability or acetylation may render H2A.Z nucleosomes sensitive to movement or ejection by chromatin remodellers and more fully expose the promoter. Finally, one outstanding question is how H2A.Z and SWR1 are targeted to promoter nucleosomes *in vivo*, with DNA-binding proteins and histone modifications probably both contributing^{21,22}.

Transcriptional noise

Transcriptional noise refers to the variability in the expression of single gene alleles in a cell population held under constant growth conditions: low-noise genes show uniform expression across the population, whereas high-noise genes show variable expression. Here, covered promoters have a much higher level of noise, a feature demonstrated by genome-scale studies¹⁸ and careful analysis of the yeast *PHO5* gene⁵⁴, a covered promoter. Covered promoters may be noisier than open promoters because of the relative differences in chromatin structure between their basal and active states⁵⁴. Open promoters have a pronounced NDR that encompasses and exposes their relatively few transcription-factor binding sites. The transition of open promoters from moderate constitutive expression to a higher active state may involve modest additional nucleosome depletion at the NDR, -1 or +1 site, and perhaps the binding of one additional factor. By contrast, many covered promoters require major changes in both nucleosome positioning and occupancy, requiring the extensive use of chromatin remodellers to permit full binding by transcription factors. Certain covered promoters may therefore be more bistable in nature — prone to bursts of transcription followed by periods of chromatin repression, rather than the steady reinitiation allowed by constitutively open promoters. Although chromatin plays important roles, there are other sources of noise in gene regulation, including signalling and transcription-factor regulation⁵⁵.

'Statistical' positioning and phasing

Most yeast genes, and many active genes in metazoans, have on their coding region 'phased' nucleosome arrays, in which the nucleosomes display consistent alignment and spacing relative to the physical map. These phased nucleosome arrays may help prevent the cryptic internal initiation of transcription within genes^{44,56,57}, which can produce dominant truncation derivatives. Phased arrays typically start at the

+1 nucleosome and decay slowly over distances, but it is not clear how they are formed. One possible mechanism is 'statistical' positioning, in which phased arrays emanate from a decisive element: a strongly positioned nucleosome or a nucleosome-free region⁵⁸. Either of these elements could establish a 'boundary' from which adjacent nucleosomes acquire phasing, resulting in a statistically positioned array that decays in phasing coherence over a distance.

As described earlier, open promoters may use the poly(dA:dT) stretch flanked by strong NPS elements at the -1 and +1 nucleosome to fix the position of these two nucleosomes, forming a decisive boundary element from which additional nucleosomes could, in principle, be statistically positioned in both directions without a requirement for additional flanking NPS elements²⁵. This phasing could involve the spacing function of ISWI-family remodellers⁴² acting to evenly space nucleosomes upstream and downstream of the positioned -1 and +1 nucleosomes, in open promoters and coding regions, respectively (Fig. 1).

Paused Pol II and nucleosome positioning

Pol II pauses at the edge of the first nucleosome downstream of the TSS at many genes in humans⁵⁹ (see page 186), and these promoters often lack a nucleosome at the -1 position (the NDR in humans). This raises the possibility that the +1 nucleosome is involved in pausing, either by physically blocking progression or by regulating the presence or activity of factors that help Pol II overcome this pause. In human cells, at inactive genes the +1 nucleosome resides ~10 bp downstream of the TSS, whereas at active genes the +1 nucleosome resides ~40 bp downstream of the TSS, suggesting that the +1 nucleosome slides as part of the transition to processive elongation¹³. The modification state of the +1 nucleosome is highly regulated; for example, it has high levels of histone H3 trimethylated at Lys 4 (H3K4me3), which binds the PHD domain in the Taf3 subunit of TFIID, and also has acetylation, which binds the Bdf1/2 subunits of TFIID⁶⁰. In principle, TFIID, the +1 nucleosome and Pol II could form a functional unit within which initiation and elongation are regulated. In humans, the phasing of nucleosomes in the coding region is largely confined to genes with active Pol II¹³. One possibility is that in humans this functional unit serves as a boundary around which nucleosomes are subsequently positioned after Pol II pauses, perhaps by statistical positioning and ISWI function, although this remains to be tested. Experiments to address the roles of remodellers in regulating the +1 nucleosome in higher eukaryotes are also eagerly awaited.

Promoter nucleosomes have high turnover

One model of inactive promoters posits that they are static entities with immobile nucleosomes. However, if this is true, it might be difficult for certain pioneer transcription factors to find their sites in promoters, especially if their sites reside within the edge of a nucleosome. Recent results in yeast have shown that promoters have 'hot' nucleosomes; that is, nucleosomes with high turnover rates in both their repressed and active states^{61,62}. Remarkably, promoter nucleosomes are hotter than those in coding regions, even at active genes with transcribing Pol II. This suggests that the intense focus of chromatin modifying and remodelling enzymes (and H2A.Z) near the future TSS greatly promotes nucleosome turnover, allowing the inspection of promoters by transcription factors at a tuned rate. Acetylation at H3K56 has been shown to be strongly correlated with hot nucleosomes⁶². One attractive possibility is that histone modifications such as H3K56ac attract SWI/SNF-family remodellers to eject nucleosomes^{61,62}. Finally, histone chaperones have been shown to assist both the replication-independent histone deposition machinery and remodellers during histone eviction^{39,63}. Indeed, the coordinated action of remodellers and chaperones is probably at the forefront of histone eviction and assembly dynamics.

Concluding remarks

Over the past five years, genome-wide and gene-specific studies have revealed important concepts for the regulation of promoter architecture (Box 1). Among the recent advances is a greater understanding of how DNA sequences help define nucleosome positions, and how chromatin

modification and remodelling machineries optimize both the initial repressed (or basal) state and the transition to the active state. Here, the notion that promoters are poised in the repressed state is becoming better appreciated, with nucleosome positioning and depletion, histone variants, histone modifications and (in metazoans) the presence of a paused Pol II all being prominent aspects of this poised state. Furthermore, these poised, repressed architectures are now understood as dynamic structures themselves, with high nucleosome turnover. Indeed, it is the precise nature of the poised state that sets the requirements for the transition to the active state. Finally, the effect of distal enhancer proteins on promoter dynamics is an important area of current and future study. The transitions of promoters between their repressed and active states have been studied for decades, but our recent ability to simultaneously monitor many aspects of promoter architecture, modification and composition at high temporal and spatial resolution will doubtless reveal welcome detail in these individual steps, and this can also be extended to enhancers. In addition, genetic experiments are required to determine the dependency relationships — to see which factors, histone modifications and nucleosome movements are truly instructing the next transition. Understanding these relationships will greatly enhance our understanding of the logic and orchestration of promoter regulation. ■

- Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128**, 707–719 (2007).
- Narlikar, G. J., Fan, H. Y. & Kingston, R. E. Cooperation between complexes that regulate chromatin structure and transcription. *Cell* **108**, 475–487 (2002).
- Becker, P. B. & Horz, W. ATP-dependent nucleosome remodeling. *Annu. Rev. Biochem.* **71**, 247–273 (2002).
- Saha, A., Wittmeyer, J. & Cairns, B. R. Chromatin remodelling: the industrial revolution of DNA around histones. *Nature Rev. Mol. Cell Biol.* **7**, 437–447 (2006).
- Lee, C. K., Shibata, Y., Rao, B., Strahl, B. D. & Lieb, J. D. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nature Genet.* **36**, 900–905 (2004).
- Lee, W. *et al.* A high-resolution atlas of nucleosome occupancy in yeast. *Nature Genet.* **39**, 1235–1244 (2007).
- Yuan, G. C. *et al.* Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**, 626–630 (2005).
This paper provides the first high-resolution view of nucleosome positioning and reveals a nucleosome-depleted region that overlaps with transcription-factor binding sites.
- Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
- Bernstein, B. E., Liu, C. L., Humphrey, E. L., Perlstein, E. O. & Schreiber, S. L. Global nucleosome occupancy in yeast. *Genome Biol.* **5**, R62 (2004).
- Albert, I. *et al.* Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature* **446**, 572–576 (2007).
- Mavrich, T. N. *et al.* Nucleosome organization in the *Drosophila* genome. *Nature* **453**, 358–362 (2008).
- Venters, B. J. & Pugh, B. F. A canonical promoter organization of the transcription machinery and its regulators in the *Saccharomyces* genome. *Genome Res.* **19**, 360–371 (2009).
- Schones, D. E. *et al.* Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887–898 (2008).
- Segal, E. *et al.* A genomic code for nucleosome positioning. *Nature* **442**, 772–778 (2006).
This influential paper provides computational prediction of nucleosome positions.
- Kaplan, N. *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**, 362–366 (2009).
- Segal, E. & Widom, J. Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.* **19**, 65–71 (2009).
- Ioshikhes, I. P., Albert, I., Zanton, S. J. & Pugh, B. F. Nucleosome positions predicted through comparative genomics. *Nature Genet.* **38**, 1210–1215 (2006).
- Tirosh, I. & Barkai, N. Two strategies for gene regulation by promoter nucleosomes. *Genome Res.* **18**, 1084–1091 (2008).
Several concepts in promoter classification developed in this paper are central to my Review.
- Field, Y. *et al.* Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput. Biol.* **4**, e1000216 (2008).
- Guillemette, B. *et al.* Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. *PLoS Biol.* **3**, e384 (2005).
- Raisner, R. M. *et al.* Histone variant H2A.Z marks the 5' ends of both active and inactive genes in eukaryotic chromatin. *Cell* **123**, 233–248 (2005).
- Zhang, H., Roberts, D. N. & Cairns, B. R. Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell* **123**, 219–231 (2005).
- Struhl, K. Naturally occurring poly(dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. *Proc. Natl Acad. Sci. USA* **82**, 8419–8423 (1985).
- Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).
- Mavrich, T. N. *et al.* A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.* **18**, 1073–1083 (2008).
- Almer, A., Rudolph, H., Hinnen, A. & Horz, W. Removal of positioned nucleosomes from the yeast *PHO5* promoter upon *PHO5* induction releases additional upstream activating

- DNA elements. *EMBO J.* **5**, 2689–2696 (1986).
27. Fascher, K. D., Schmitz, J. & Horz, W. Role of *trans*-activating proteins in the generation of active chromatin at the *PHO5* promoter in *S. cerevisiae*. *EMBO J.* **9**, 2523–2528 (1990).
 28. Lam, F. H., Steger, D. J. & O'Shea, E. K. Chromatin decouples promoter threshold from dynamic range. *Nature* **453**, 246–250 (2008).
 29. Hebbar, P. B. & Archer, T. K. Chromatin remodeling by nuclear receptors. *Chromosoma* **111**, 495–504 (2003).
 30. Basehoar, A. D., Zanton, S. J. & Pugh, B. F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**, 699–709 (2004).
 31. Huisinga, K. L. & Pugh, B. F. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in *Saccharomyces cerevisiae*. *Mol. Cell* **13**, 573–585 (2004).
 32. Martinez-Campa, C. *et al.* Precise nucleosome positioning and the TATA box dictate requirements for the histone H4 tail and the bromodomain factor Bdf1. *Mol. Cell* **15**, 69–81 (2004).
 33. Grant, P. A., Sterner, D. E., Duggan, L. J., Workman, J. L. & Berger, S. L. The SAGA unfolds: convergence of transcription regulators in chromatin-modifying complexes. *Trends Cell Biol.* **8**, 193–197 (1998).
 34. Kingston, R. E. & Narlikar, G. J. ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. *Genes Dev.* **13**, 2339–2352 (1999).
 35. Owen-Hughes, T. Colworth memorial lecture. Pathways for remodelling chromatin. *Biochem Soc. Trans.* **31**, 893–905 (2003).
 36. Langst, G., Bonte, E. J., Corona, D. F. & Becker, P. B. Nucleosome movement by CHRAC and ISWI without disruption or *trans*-displacement of the histone octamer. *Cell* **97**, 843–852 (1999).
 37. Ito, T., Bulger, M., Pazin, M. J., Kobayashi, R. & Kadonaga, J. T. ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell* **90**, 145–155 (1997).
 38. Whitehouse, I. *et al.* Nucleosome mobilization catalysed by the yeast SWI/SNF complex. *Nature* **400**, 784–787 (1999).
 39. Lorch, Y., Zhang, M. & Kornberg, R. D. Histone octamer transfer by a chromatin-remodeling complex. *Cell* **96**, 389–392 (1999).
 - This paper shows that a chromatin remodeller can eject a nucleosome.**
 40. Mizuguchi, G. *et al.* ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* **303**, 343–348 (2004).
 - This paper provides the first evidence for *in vitro* insertion of H2A.Z by a remodeller.**
 41. Corona, D. F., Clapier, C. R., Becker, P. B. & Tamkun, J. W. Modulation of ISWI function by site-specific histone acetylation. *EMBO Rep.* **3**, 242–247 (2002).
 42. Kagalwala, M. N., Glaus, B. J., Dang, W., Zofall, M. & Bartholomew, B. Topography of the ISW2-nucleosome complex: insights into nucleosome spacing and chromatin remodeling. *EMBO J.* **23**, 2092–2104 (2004).
 43. Gelbart, M. E., Bachman, N., Delrow, J., Boeke, J. D. & Tsukiyama, T. Genome-wide identification of Isw2 chromatin-remodeling targets by localization of a catalytically inactive mutant. *Genes Dev.* **19**, 942–954 (2005).
 44. Whitehouse, I., Rando, O. J., Delrow, J. & Tsukiyama, T. Chromatin remodelling at promoters suppresses antisense transcription. *Nature* **450**, 1031–1035 (2007).
 45. Whitehouse, I. & Tsukiyama, T. Antagonistic forces that position nucleosomes *in vivo*. *Nature Struct. Mol. Biol.* **13**, 633–640 (2006).
 46. Parnell, T. J., Huff, J. T. & Cairns, B. R. RSC regulates nucleosome positioning at Pol II genes and density at Pol III genes. *EMBO J.* **27**, 100–110 (2008).
 47. Badis, G. *et al.* A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Mol. Cell* **32**, 878–887 (2008).
 48. Hartley, P. D. & Madhani, H. D. Mechanisms that specify promoter nucleosomes and identity. *Cell* **137**, 445–458 (2009).
 49. Henikoff, S. Nucleosome destabilization in the epigenetic regulation of gene expression. *Nature Rev. Genet.* **9**, 15–26 (2008).
 50. Keogh, M. C. *et al.* The *Saccharomyces cerevisiae* histone H2A variant Htz1 is acetylated by NuA4. *Genes Dev.* **20**, 660–665 (2006).
 51. Millar, C. B., Xu, F., Zhang, K. & Grunstein, M. Acetylation of H2AZ Lys 14 is associated with genome-wide gene activity in yeast. *Genes Dev.* **20**, 711–722 (2006).
 52. Jin, C. & Felsenfeld, G. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev.* **21**, 1519–1529 (2007).
 53. Ahmad, K. & Henikoff, S. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell* **9**, 1191–1200 (2002).
 54. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814 (2004).
 - This important paper connects chromatin attributes to transcriptional noise.**
 55. Raser, J. M. & O'Shea, E. K. Noise in gene expression: origins, consequences, and control. *Science* **309**, 2010–2013 (2005).
 56. Kaplan, C. D., Laprade, L. & Winston, F. Transcription elongation factors repress transcription initiation from cryptic sites. *Science* **301**, 1096–1099 (2003).
 57. Carrozza, M. J. *et al.* Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**, 581–592 (2005).
 58. Kornberg, R. D. & Stryer, L. Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic Acids Res.* **16**, 6677–6690 (1988).
 59. Core, L. J. & Lis, J. T. Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science* **319**, 1791–1792 (2008).
 60. Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. *Nature Rev. Genet.* **10**, 161–172 (2009).
 61. Dion, M. F. *et al.* Dynamics of replication-independent histone turnover in budding yeast. *Science* **315**, 1405–1408 (2007).
 62. Rufiange, A., Jacques, P. E., Bhat, W., Robert, F. & Nourani, A. Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1. *Mol. Cell* **27**, 393–405 (2007).
 63. Adkins, M. W., Howar, S. R. & Tyler, J. K. Chromatin disassembly mediated by the histone chaperone Asf1 is essential for transcriptional activation of the yeast *PHO5* and *PHO8* genes. *Mol. Cell* **14**, 657–666 (2004).

Acknowledgements I thank T. Parnell and C. Clapier for comments and assistance with figures. I am grateful for support from the US National Institutes of Health (grant GM60415) and the Howard Hughes Medical Institute.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Correspondence should be addressed to B.R.C. (brad.cairns@hci.utah.edu).

Genomic views of distant-acting enhancers

Axel Visel^{1,2}, Edward M. Rubin^{1,2} & Len A. Pennacchio^{1,2}

In contrast to protein-coding sequences, the significance of variation in non-coding DNA in human disease has been minimally explored. A great number of recent genome-wide association studies suggest that non-coding variation is a significant risk factor for common disorders, but the mechanisms by which this variation contributes to disease remain largely obscure. Distant-acting transcriptional enhancers — a major category of functional non-coding DNA — are involved in many developmental and disease-relevant processes. Genome-wide approaches to their discovery and functional characterization are now available and provide a growing knowledge base for the systematic exploration of their role in human biology and disease susceptibility.

Multiple lines of evidence indicate that important functional properties are embedded in the non-coding portion of the human genome, but identifying and defining these features remains a major challenge. An initial estimate of the magnitude of functional non-coding DNA was derived from comparative analysis of the first available mammalian genomes (human and mouse), which indicated that fewer than half of the evolutionary constrained sequences in the human genome encode proteins¹, a prospect that gained further support when additional vertebrate genomes became available for comparative genomic analyses².

The overall impact of these presumably functional non-coding sequences on human biology was initially unclear. A considerable urgency to define their locations and functions came from a growing number of known associations of non-coding sequence variants with common human diseases. Specifically, genome-wide association studies (GWAS) have revealed a large number of disease susceptibility regions that do not overlap protein-coding genes but rather map to non-coding intervals. For example, a 58-kilobase linkage disequilibrium block located at human chromosome 9p21 was shown to be reproducibly associated with an increased risk for coronary artery disease, yet the risk interval lies more than 60 kilobases away from the nearest known protein-coding gene^{3,4}. To estimate the global contribution of variation in non-coding sequences to phenotypic and disease traits, we performed a meta-analysis of ~1,200 single-nucleotide polymorphisms (SNPs) identified as the most significantly associated variants in GWAS published so far (ref. 5, accessed 2 March 2009). Using conservative parameters that tend to overestimate the size of linkage disequilibrium blocks, we found that in 40% of cases (472 of 1,170) no known exons overlap either the linked SNP or its associated haplotype block, suggesting that in more than one-third of cases non-coding sequence variation causally contributes to the traits under investigation.

One possibility that could explain these GWAS hits is that the non-coding intervals contain enhancers, a category of gene regulatory sequence that can act over long distances. A simplified view of the current understanding of the role of enhancers in regulating genes is summarized in Fig. 1. The docking of RNA polymerase II to proximal promoter sequences and transcription initiation are fairly well characterized; by contrast, the mechanisms by which insulator and silencer elements buffer or repress gene regulation, respectively, are less well understood⁶. Transcriptional enhancers are regulatory sequences that can be located upstream of, downstream of or within their target gene and can modulate expression independently of their orientation⁷. In vertebrates, enhancer sequences are thought to comprise densely clustered aggregations of transcription-factor-binding

sites⁸. When appropriate occupancy of transcription-factor-binding sites is achieved, recruitment of transcriptional coactivators and chromatin-remodelling proteins occurs. The resultant protein aggregates are thought to facilitate DNA looping and ultimately promoter-mediated gene activation (see page 199). In-depth studies of individual genes such as *APOE* or *NKX2-5* (reviewed in ref. 9) have shown that many genes are regulated by complex arrays of enhancers, each driving distinct aspects of the messenger RNA expression pattern. These modular properties of mammalian enhancers are also supported by their additive regulatory activities in heterologous recombination experiments¹⁰.

The purely genetic evidence from GWAS does not allow any direct inferences regarding the underlying molecular mechanisms, but a number of in-depth studies of individual loci (see below) suggest that variation in distant-acting enhancer sequences and the resultant changes in their activities can contribute to human disorders. Although we anticipate a variety of other non-coding functional categories such as negative gene regulators or non-coding RNAs to have a role in human disease, in this Review we focus on the role of enhancers and on strategies to define their location and function throughout the genome.

Enhancers in human disease

Beginning with the discovery that an inherited change in the β -globin gene alters one of the coded amino acids and thereby causes sickle-cell anaemia^{11,12}, thousands of mutations in the coding regions of genes have been identified to be responsible for monogenic disorders over the past half century. By contrast, the role of mutations not involving primary gene structural sequences has been minimally explored, largely owing to our inability to recognize relevant non-coding sequences, much less predict their function. The molecular genetic identification of individual enhancers involved in disease has been, in most cases, a painstaking and inefficient endeavour. Nevertheless, a number of successful studies have shown that distant-acting gene enhancers exist in the human genome and that variation in their sequences can contribute to disease. In this section, we discuss three examples in which enhancers were directly shown to play a role in human disease: thalassaemias resulting from deletions or rearrangements of β -globin gene (*HBB*) enhancers, preaxial polydactyly resulting from sonic hedgehog (*SHH*) limb-enhancer point mutations, and susceptibility to Hirschsprung's disease associated with a *RET* proto-oncogene enhancer variant.

The extensive studies of the human globin system and its role in haemoglobinopathies have historically served as a test bed for defining not only the role of coding sequences in disease^{11,12} but also that of non-coding

¹Genomics Division, MS 84-171, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ²US Department of Energy Joint Genome Institute, Walnut Creek, California 94598, USA.

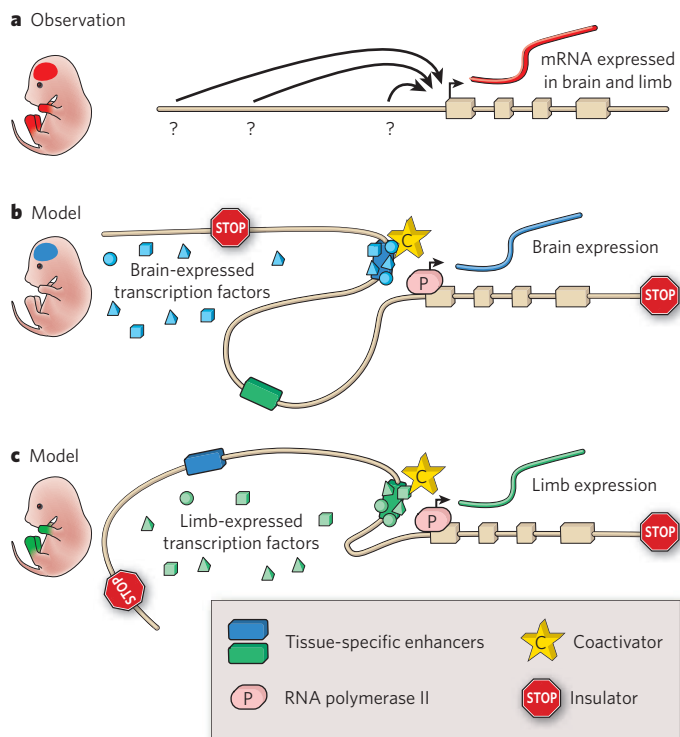


Figure 1 | Overview of gene regulation by distant-acting enhancers.

a, For many genes, the regulatory information embedded in the promoter is insufficient to drive the complex expression pattern observed at the messenger RNA level. For example, a gene could be expressed both in the brain and in the limbs during embryonic development (red), even if the promoter by itself is not active in either of these structures, suggesting that appropriate expression depends on additional sequences that are distant-acting and *cis*-regulatory. However, defining the genomic locations of such regulatory elements (question marks) and their activities in time and space (arrows) is a major challenge. **b, c**, Tissue-specific enhancers are thought to contain combinations of binding sites for different transcription factors. Only when all required transcription factors are present in a tissue does the enhancer become active: it binds to transcriptional coactivators, relocates into physical proximity with the gene promoter (through a looping mechanism) and activates transcription by RNA polymerase II. In any given tissue, only a subset of enhancers is active, as schematically shown in **b** and **c** for the example gene pictured in **a**, whose expression is controlled by two separate enhancers with brain-specific and limb-specific activities. Insulator elements prevent enhancer–promoter interactions and can thus restrict the activity of enhancers to defined chromatin domains. In addition to activation by enhancers, negative regulatory elements (including repressors and silencers) can contribute to transcriptional regulation (not shown).

sequences. The α -thalassaemias and β -thalassaemias are haemoglobinopathies resulting from imbalances in the ratio of α -globin to β -globin chains in red blood cells. The molecular basis of these conditions was initially elucidated in cases in which inactivation or deletion of globin structural genes could be readily identified¹³. However, although gene deletion or sequence changes resulting in a truncated or non-functional gene product explained some thalassaemia cases, for a subset of patients intensive sequencing efforts failed to reveal abnormalities in globin protein-coding sequences. Through extensive long-range mapping and sequencing of DNA from individuals diagnosed with thalassaemia but lacking globin coding mutations, it was eventually discovered that many of these globin chain imbalances were due to deletion or chromosome rearrangements that resulted in the repositioning of distant-acting enhancers required for normal globin gene expression^{14,15}. These early molecular genetic studies revealed a clear role for non-coding regulatory elements as a cause of human disorders through their impact on gene expression. Since then, many such examples of ‘position effects’, defined as changes in the expression of a gene when its location in a chromosome is changed, often by translocation, have been found¹⁶.

In addition to the pathological consequences of the removal or the repositioning of distant-acting enhancers, there are also examples of single-nucleotide changes within enhancer elements as a cause of human disorders. One example of this category of disease-causing non-coding mutation involves the limb-specific long-distance enhancer ZRS (also known as MFCS1) of *SHH* (Fig. 2). This enhancer is located at the extreme distance of approximately 1 megabase from *SHH*, within the intron of a neighbouring gene^{17,18}. Of interest is that, initially, the gene in which the enhancer resides was thought to be relevant for limb development and was therefore named limb region 1 (*LMBR1*)¹⁹. Facilitated by the functional knowledge of the ZRS enhancer from mouse studies, targeted resequencing screens of this enhancer in humans revealed that it is associated with preaxial polydactyly. Approximately a dozen different single-nucleotide variations in this regulatory element have been identified in humans with preaxial polydactyly and segregate with the limb abnormality in families^{18,20}. Studies of the impact of the human ZRS sequence changes have been carried out in transgenic mice, in which the single-nucleotide changes result in ectopic anterior-limb expression during development, consistent with preaxial digit outgrowth²¹. Furthermore, sequence changes in the orthologous enhancers were found in mice, as well as in cats, with preaxial polydactyly^{22,23}, and targeted deletion of the enhancer in mice caused truncation of limbs¹⁷. These studies illustrate the importance of first experimentally identifying distant-acting enhancers in allowing subsequent human genetic studies to explore the potential role of disease-causing mutation in functional non-coding sequences.

Another example of enhancer variation contributing to human disease is provided by the discovery of a common non-coding variant linked to susceptibility to Hirschsprung’s disease. Although multigenic, Hirschsprung’s disease risk is strongly linked to coding mutations in the *RET* proto-oncogene^{24,25}. However, family-based studies have also revealed evidence for Hirschsprung’s disease linked to the *RET* locus in people lacking any accompanying functional *RET* coding mutations. Through the use of multispecies comparisons of orthologous genomic intervals that include and flank *RET*, coupled with *in vitro* and *in vivo* functional studies, an enhancer sequence located in intron 1 of *RET* was identified and found to contain a common variant contributing more than a 20-fold increased risk for Hirschsprung’s disease than rarer alleles in this element^{26,27}. In transgenic mice, this enhancer was shown to be active in the nervous system and digestive tract during embryogenesis in a manner consistent with its putative role in Hirschsprung’s disease²⁷. It is interesting to note that although this enhancer variation is clearly important in disease risk, the variant alone is not sufficient to cause Hirschsprung’s disease, highlighting the complex aetiology of this disorder.

As is evident from these labour-intensive gene-centric studies, enhancers can, in principle, have an important role in disease, but it remains unclear whether these are rare exceptions or whether variation in enhancers contributes to disease on a pervasive scale. Support for the latter comes from a rapidly growing number of examples in which non-coding SNPs linked to disease traits through GWAS were found to affect the expression levels of nearby genes²⁸, suggesting that variation in regulatory sequences may commonly contribute to a wide range of disorders. The results of the recent GWAS, coupled with the role of gene regulation in normal human biology, provide a strong incentive for defining the distant-acting-enhancer architecture of the human genome.

Harnessing evolution

Gene-centric studies have been crucial to defining the general characteristics of gene regulatory regions in specific human disorders, but they have only identified and characterized a limited number of such elements. Systematic large-scale identification of sequences that are likely to be enhancers was first made possible by comparative genomic strategies. These approaches are based on the assumption that the sequences of gene regulatory elements, like those of protein-coding genes, are under negative evolutionary selection, because most changes in functional sequences have deleterious consequences^{29–32}. Thus, it was proposed that statistical measures of evolutionary sequence constraint would provide a way to

identify potential enhancer sequences within the vast amount of non-coding sequence in the human genome. Support for this approach initially came from retrospective comparative genomic analyses of experimentally well-defined enhancers; these analyses revealed that enhancers frequently shared sequence conservation with orthologous regions present in the genomes of other mammals. The observation that DNA conservation identified many of these complex regulatory elements encouraged investigators to move away from blind studies of regions flanking genes of interest towards focusing specifically on non-coding sequences constrained across vertebrate species, culminating in whole-genome studies in which conservation level alone guided experimentation^{32–34}.

Initially, comparisons over extreme evolutionary distances, such as between humans and fish, were deemed most effective for this purpose^{29,31}. Indeed, it was observed through large-scale transgenic mouse and fish studies that many of these non-coding sequences that had been conserved for hundreds of millions of years of evolution were enhancers that drove expression in highly specific anatomical structures during embryonic development. Likewise, so-called ultraconserved non-coding elements, which are blocks of 200 base pairs or more that are perfectly conserved between humans, mice and rats³⁵, were also found to be highly enriched in tissue-specific enhancers, suggesting that the success rate of comparative approaches for enhancer identification depends on scoring criteria, rather than just evolutionary distance³². This idea was further supported by the development of advanced statistical tools designed to quantify evolutionary constraint, from which it became evident that even comparisons between relatively closely related species can be effective predictors of enhancers^{2,36,37}. A large-scale transgenic mouse study that included nearly all non-exonic ultraconserved elements in the human genome revealed that whereas many of them are developmental *in vivo* enhancers, other conserved non-coding sequences that are under similar evolutionary constraint, but are not perfectly conserved between humans and mice, are equally enriched in enhancers³³. These results suggest that ultraconserved elements do not represent a functionally distinct subgroup of conserved non-coding sequences in terms of their enrichment in *in vivo* enhancers but rather that there is a much larger number of non-coding sequences that are under similar evolutionary constraint and are just as enriched in enhancers as are ultraconserved elements.

Independent of the specific algorithms and metrics that were used, most categories of conserved non-coding sequence were found not to be randomly distributed in the genome. Instead, they are located in a highly

biased manner near genes active during development^{2,33–35}, consistent with the observation that a large proportion of these non-coding sequences give robust positive signals in various assays of being tissue-specific *in vivo* enhancers active during development.

Comparative approaches are an effective high-throughput genomic strategy for identifying non-coding sequences that are highly likely to be enhancers, but they have several limitations. First, although conservation is indicative of function, it is not necessarily indicative of enhancer activity, because many other types of non-coding functional element that may have similar conservation signatures are known to exist. Second, even when conservation of non-coding DNA results from enhancer function, conservation cannot predict when and where an enhancer is active in the developing or adult organism. For all identified candidates, experimental studies are needed to decipher the gene-regulatory properties of each element, and these transgenic studies cannot feasibly be scaled to generate truly comprehensive genome-wide data sets.

A perplexing study questioning the importance of extremely conserved enhancers found the lack of an apparent phenotype upon targeted deletion of four independent ultraconserved elements in mice³⁸. General expectations were that non-coding sequences that have been perfectly conserved in mammals for tens of millions of years must be essential and that their deletion should result in severe phenotypes, comparable to those observed upon deletion of the *Shh* limb enhancer and other less well-conserved enhancers^{9,17}. However, mice with deletions of such ultraconserved enhancers were viable, fertile and showed no overt phenotype³⁸. Interpretations of this lack of obvious effect are similar to those of the absence of phenotypes upon deletion of highly conserved protein-coding genes: minor phenotypes may have escaped detection in the assays used; there may have been functional redundancy with other genes or enhancers; or there may have been reductions in fitness that only become apparent over multiple generations or are not easily detected in a controlled laboratory environment. This study highlighted that although extreme non-coding sequence conservation is an effective predictor of the location of enhancers in the genome, the degree of evolutionary constraint is not directly correlated with the severity of anticipated phenotypes.

Sequencing-based enhancer discovery

As a strategy complementary to comparative genomic methods, it has recently become possible to generate genome-wide maps of chromatin marks that can be used to identify the location of enhancers and other

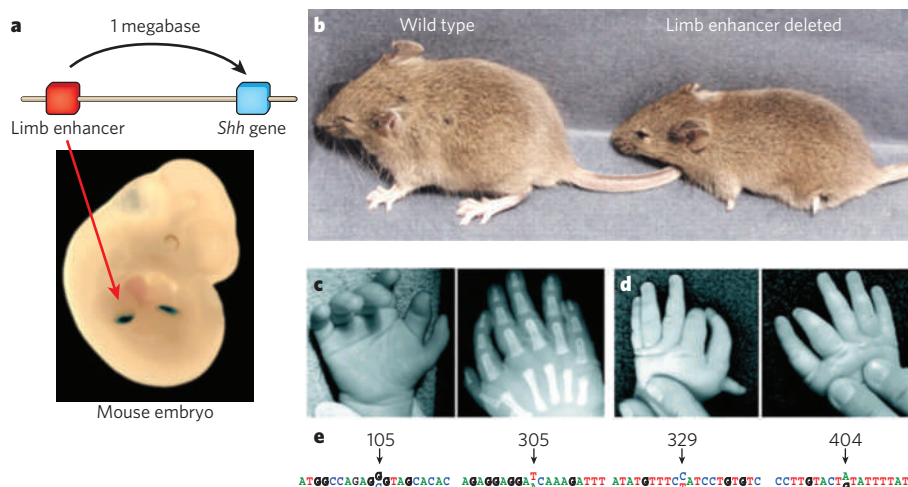


Figure 2 | Consequences of deletion and mutation of the limb enhancer of sonic hedgehog. **a**, The limb enhancer of *Shh* is located approximately 1 megabase away from its target promoter in the intron of a neighbouring gene (*Lmbr1*; exons not shown). In transgenic mouse reporter assays, this non-coding sequence targets gene expression to a posterior region of the developing limb bud (red arrow). (Image reproduced, with permission, from ref. 18.) **b**, Mice with a targeted deletion of this enhancer have severely truncated limbs, which strikingly demonstrates its functional importance

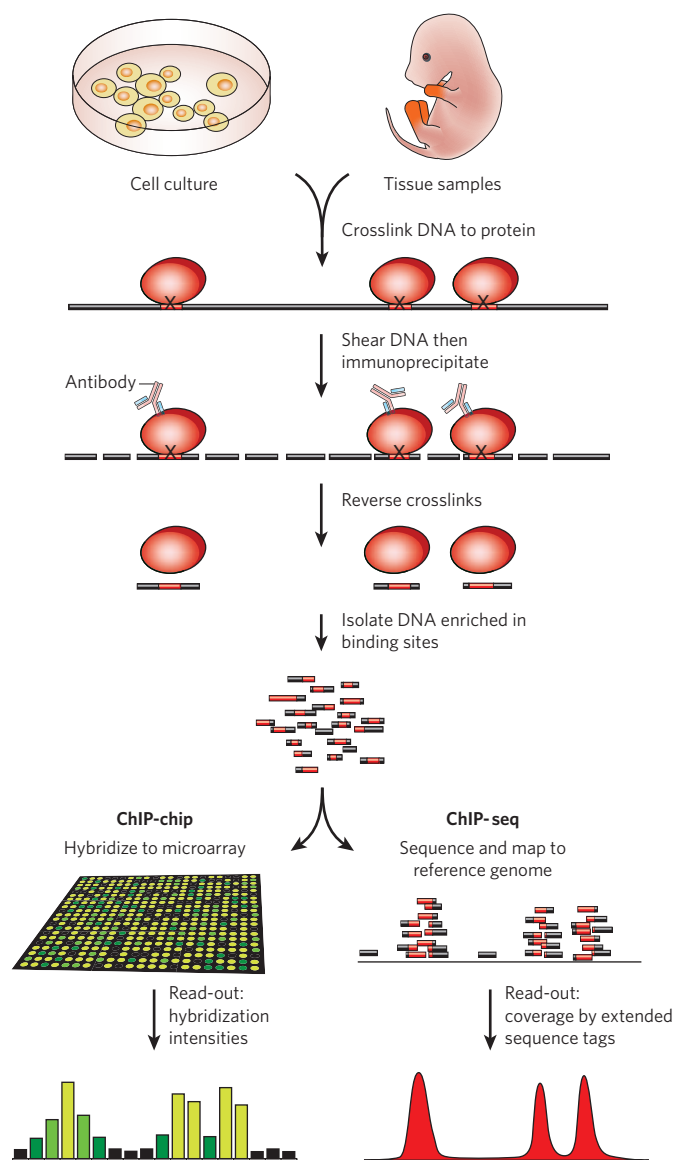
in development. (Reproduced, with permission, from ref. 17.) **c–e**, Point mutations in the orthologous human enhancer sequence result in preaxial polydactyly, emphasizing the potential significance of variation in non-coding functional sequences in both rare and common human disorders: **c** and **d** show the hands of two patients with point mutations in the *SHH* limb enhancer; **e** shows point mutations associated with preaxial polydactyly identified in four unrelated families. (Panels **c** and **d** reproduced, with permission, from ref. 18; panel **e** modified, with permission, from ref. 18.)

Box 1 | Mapping of regulatory elements using ChIP-chip and ChIP-seq

Formaldehyde crosslinking of DNA to proteins that bind to it directly or as part of larger complexes⁷¹, combined with subsequent immunoprecipitation targeting specific DNA-associated proteins (ChIP⁷²), was widely used in the pre-genomic era to study protein–DNA interactions directly in cultured cells or in tissue samples. The top portion of the figure shows a schematic overview of the individual steps involved. They include the molecular fixation of non-covalent protein–DNA interactions, shearing of the crosslinked chromatin, immunoprecipitation with an antibody binding the protein of interest and reversal of crosslinks. In many cases, antibodies that bind to covalently modified proteins are used, for example those that recognize methyl groups at defined amino-acid residues of histones. In the conventional ChIP approach, enrichment of the associated DNA fragments relative to non-immunoprecipitated (‘input’) DNA is quantified for individual proposed binding locations (not shown). This need for quantification at every site of interest initially thwarted the application of ChIP on a genomic scale.

The introduction of DNA microarrays allowed the hybridization-based interrogation of large numbers of potential binding sites in parallel (ChIP-on-chip, or ChIP-chip), thus making it possible to screen entire compact model-organism genomes^{73,74} or large vertebrate genome intervals⁷⁵ in a single experiment (see figure, bottom left). ChIP-chip was used on a massive scale in the Encyclopedia of DNA Elements (ENCODE) pilot project, in which dozens of proteins and protein modifications were initially mapped in a representative 1% portion of the human genome³⁹.

Recently, chromatin immunoprecipitation coupled to massively parallel sequencing (ChIP-seq) has become increasingly used as an alternative to ChIP-chip^{44–47}. The ChIP-seq method is very similar to the experimental set-up of ChIP-chip, except that, in the final step, next-generation sequencing techniques are used to determine the sequence of immunoprecipitated DNA fragments, which are then computationally mapped to the reference genome (see figure, bottom right). Improved sequencing technologies offer the possibility to obtain millions of mappable reads in a single ChIP-seq experiment at moderate cost. The results from ChIP-seq are based on statistical analysis of read counts, which overcomes many of the challenges associated with the quantification and normalization of hybridization signals, and an increasing number of advanced computational ChIP-seq analysis tools are becoming available⁷⁶. ChIP-seq analysis covers by default the entire mappable portion of the reference genome without the need to restrict the analysis to its subregions.



regulatory regions. These genomic approaches have become possible as a result of an improved understanding of the proteins and epigenetic marks found at particular categories of regulatory element, together with concurrently developed technologies that allow traditional chromatin immunoprecipitation (ChIP) techniques to be applied on the scale of whole vertebrate genomes. The initial in-depth studies of 1% of the genome in the Encyclopedia of DNA Elements (ENCODE) pilot project³⁹ were largely based on data sets generated by the ChIP-chip technique (Box 1) and revealed the molecular properties of a variety of regulatory elements.

With respect to enhancer identification, a particularly relevant insight was the identification of specific histone methylation signatures found at enhancers. In contrast to promoters, which are marked by trimethylation of histone H3 at lysine residue 4 (H3K4me3), active enhancers are marked by monomethylation at this position (H3K4me1)⁴⁰. Mapping these marks in the ENCODE regions and, more recently, throughout the entire genome⁴¹ revealed tens of thousands of elements that were predicted to be active enhancers in the examined cell types. Importantly, these predicted enhancers were also frequently associated with the transcriptional coactivators p300 and/or TRAP220 (also known as MED1), raising the possibility that such coactivators might be useful general markers for mapping enhancers. Although it was initially not

clear to what extent the presence of transcriptional coactivators such as p300 is indicative of active rather than inactive enhancers, comparison of DNase I hypersensitivity (a marker of open chromatin structure) in several cell lines throughout the ENCODE regions revealed that the location of cell-line-specific distal DNase-I-hypersensitivity sites correlates with cell-line-specific p300 binding at these sites, providing further support for the possibility that transcriptional coactivators, along with histone modification signatures, may be useful for the mapping of DNA elements with cell-specific and tissue-specific enhancer activities⁴².

Owing to the development of the ChIP-seq technique (Box 1), which has now superseded ChIP-chip as the method of choice for many applications, genome-wide maps for a considerable number of chromatin marks and transcription factors both in humans and mice have become available^{43–55}. These data sets allowed the identification of not only the H3K4me1 and H3K4me3 signatures discussed earlier but also additional chromatin marks present at predicted or validated enhancers, and provided a refined view of their correlation to enhancer activities^{44,51,55}. However, with very few exceptions (see, for example, refs 50 and 54) genome-wide mapping of these and other regulation-associated chromatin marks (Table 1) was done in immortalized cell lines, cultured stem cells or primary cell cultures. Thus, the maps of potentially enhancer-associated marks produced by these studies provided limited insight into

their *in vivo* distribution during embryonic development and in adult organs, most probably concealing the genomic location of enhancers that are inactive in these cells.

In a recent ChIP-seq study targeted at the prediction of enhancers that are active in a particular tissue during embryonic development, the transcriptional coactivator p300 was mapped in chromatin directly derived from embryonic mouse tissues, including the forebrain, the midbrain and the limb buds⁵⁶. Overall, several thousand p300 peaks were identified from these three tissues, with the vast majority of genome regions only being significantly enriched in one of the three tissues and located in non-coding regions distal from known promoters. Transgenic mouse experiments with almost 100 of these sequences revealed that they are developmental enhancers in almost all cases. More importantly, the tissue-specific occupancy by p300 as identified by ChIP-seq could in most cases also accurately predict the *in vivo* patterns of expression driven by these enhancers, providing an important advantage over comparative genomic methods for enhancer identification. The study also showed global enrichment in tissue-specific p300 peaks near genes that are expressed in the same tissue, again consistent with the proposed function of these genomic regions as active transcriptional enhancers.

These experimentally predicted genome-wide sets of *in vivo* enhancers also made it possible to address the controversial issue of the extent to which evolutionary conservation is a hallmark of *in vivo* enhancers⁵⁷. Several studies have shown that highly conserved non-coding elements are enriched in developmental *in vivo* enhancers^{32–34}. However, some observations have challenged such a generalized correlation between sequence conservation and enhancer activity: experimental analysis of individual loci suggested that a large proportion of enhancers cannot be detected by comparative genomics⁵⁸; the molecular marks of a surprisingly large proportion of sequences in the ENCODE regions suggested that regulatory functions are not, or are only weakly, conserved⁵⁹; and histone methylation present at orthologous loci in humans and mice did not correlate with overall increased levels of sequence conservation⁵⁹. In contrast to these findings, approximately 90% of the tissue-specific p300 peaks identified by ChIP-seq in developing mouse tissues overlapped regions that are under detectable evolutionary constraint⁵⁶. There may be variation in the degree of evolutionary constraint of enhancers that are active in different types of cell or developing tissue, but these data suggest that developmental enhancers that can be identified through p300 binding are commonly evolutionarily constrained.

Although preliminary, the selected studies reviewed here highlight the clear potential of mapping various chromatin marks for identifying and predicting the activity of transcriptional enhancers on a genome-wide scale. The continued progress in throughput increase and the cost reductions of next-generation sequencing technologies offer an increasingly powerful genome-wide means of identifying specific DNA–protein interactions. We anticipate that high-resolution genome-wide *in vivo* maps of chromatin marks will become available for comprehensive series of developing and adult tissues in normal states, as well as diseased states, providing multilayered *in vivo* annotations of the non-coding portion of our genome. It is important to realize that, despite this expected progress, we will continue to need parallel *in vitro* and *in vivo* biological studies to understand the functions associated with chromatin marks and to study conclusively the mechanisms by which sequence variation in distant-acting enhancers contributes to disease.

Defining the targets

The methods described here have considerably improved our ability to identify enhancers and their associated activity patterns on a genomic scale, but a remaining important challenge is to determine the relationships between enhancers and genes. Comparing ChIP-chip or ChIP-seq data with transcriptome data from microarrays or RNA-seq⁶⁰ can provide highly suggestive clues to the identity of the target gene of a given enhancer in a given tissue, but such comparisons do not provide the direct evidence for enhancer–promoter interactions that would be desirable in mapping tissue-specific regulatory networks on a genomic scale.

Early circumstantial evidence suggested that long-distance regulation of genes by enhancers occurs through the formation of physical chromatin loops, but it only became possible to study such interactions systematically through the introduction of the chromosome conformation capture (3C) assay and its derivative technologies⁶¹. Similar to ChIP, the 3C approach relies on formaldehyde crosslinking to capture DNA–DNA interactions directly in intact cells or cell nuclei. Previously suggested pairs of interacting sites are subsequently tested and validated one by one through the quantification of crosslinking events. In one of many examples demonstrating the utility of 3C in the analysis of distant-acting vertebrate enhancers, this technique was recently used⁶² to study chromatin interactions at the *Shh* locus, whose role in limb development was discussed in detail earlier. Using the 3C technique, it was demonstrated that the limb-specific long-range enhancer located in an intron of the *Lmbr1* gene directly interacts with increased frequency with the *Shh* promoter in limb buds but not in other tissues tested, providing important mechanistic support for its proposed role in *Shh* gene regulation in limb development. As an alternative approach to 3C, RNA tagging and recovery of associated proteins (RNA TRAP) can also be used to establish physical proximity between distal non-coding sequences and actively transcribed genes; this was first demonstrated in the mouse β -globin gene locus⁶³.

This work and other gene-centric studies (for more examples, see refs 64 and 65) were critical in shaping our understanding of enhancer–promoter interactions. However, they have the fundamental limitation that only one or very few previously proposed interactions between specific loci can be assayed per experiment. This limitation was partly overcome through the use of microarrays to analyse entire 3C libraries (called chromosome conformation capture-on-chip⁶⁶ and circular chromosome conformation capture⁶⁷, both known as 4C). By applying this approach to fetal liver and brain, it was demonstrated that the β -globin gene locus control region (LCR) makes reproducible tissue-specific contacts with other loci predominantly located on the same chromosome but in some cases dozens of megabases away from the LCR⁶⁶. Of possible relevance to the adoption of this approach for enhancer discovery is that reproducible interactions with other chromosome regions were also observed in the brain, where the LCR is thought to be inactive.

The 4C approaches are a significant improvement, but they still preclude the generation of truly genome-wide interaction networks because each experiment only reveals the genome-wide interactions of a single site of interest. This problem is partly alleviated by the chromosome conformation capture carbon copy (5C) method⁶⁸, in which a complex 3C library generated through multiplexed PCR is analysed by large-scale sequencing to generate a comprehensive ‘many-to-many’ interaction map of DNA–DNA interactions. However, owing to the need for specific

Table 1 | Selected major categories of non-coding functional element

Category	Function	Selected associated chromatin marks*
Promoter	Region that is located immediately upstream of a protein-coding gene, and binds to RNA polymerase II; where transcription is initiated	RNA polymerase II ⁴⁴ , H3K4me3 (ref. 40) (active promoters)
Enhancer	Region that activates transcription, often in a temporally and spatially restricted manner, by acting on a promoter. Enhancers can be located far from target promoters and are orientation independent	p300 (refs 40, 56), H3K4me1 (ref. 40)
Insulator	Separates active from inactive chromatin domains and interferes with enhancer activity when placed between an enhancer and promoter	CTCF ^{44,53}
Repressor/silencer	Negative regulators of gene expression	REST ⁴⁵ , SUZ12 (refs 69, 70)

*Many additional chromatin marks were found to correlate with one or several of these categories of regulatory element. Detailed descriptions of these markers and their respective binding characteristics at different types of regulatory sequence element can be found in refs 40, 41, 44, 51 and 55.

primers for each possible interacting fragment and the sequencing depth required for analysis of the resultant libraries, the application of 5C has so far been restricted to the in-depth analysis of single loci or chromosome regions.

As an alternative genome-wide approach, antibody-based methods might be used to restrict the analysis space in which DNA–DNA interactions are studied to a size that can be affordably analysed using currently available sequencing technologies. One possibility is to couple a chromatin–interaction paired-end tag (ChIA–PET) sequencing strategy to a ChIP step that enriches for chromatin fragments bound to a specific transcription factor or other chromatin mark of interest⁶⁴. Although the technical feasibility of this approach remains to be demonstrated, it has remarkable potential for enhancer discovery. This is because its application to general enhancer-associated marks such as p300 or histone methylation^{40,56} might identify, in a single step, enhancers active in a tissue of interest, as well as their respective target genes.

Perspective

Genetic and medical resequencing studies have been advanced by knowledge about the structure of protein-coding genes and a detailed understanding of the relationship between mRNA sequences and the primary structures of the proteins they encode. Through such studies, disease links have been established for a sizeable proportion of the ~20,000 protein-coding genes in the human genome. By contrast, a very limited number of changes in gene regulatory sequences have so far been linked to human disease. Consequently, an important motivation for functionally annotating the non-coding portion of the human genome and the *cis*-regulatory elements that it contains is to assess the relationship between variations in non-coding sequences and human disease. In the absence of genome-wide catalogues of functionally annotated regulatory elements, how these elements impact on human biology, as well as disease, will remain an untested hypothesis.

Despite advances in relevant technologies, functionally characterizing the distant-acting-enhancer architecture of the human genome in its entirety will be an enormous undertaking, owing to the great number of data points needed, which include dozens of tissues and cell types, as well as developmental states and possibly disease states.

A further challenge will be to link distant-acting enhancers to the genes they regulate. Linking enhancers to their cognate gene will allow the further assignment of these functional sequences to their basic ‘gene’ unit of heredity, for collective resequencing analysis.

Although we have focused on distant-acting enhancers here, there are other categories of functional element in the non-coding portion of the genome (for example insulators, negative regulators, promoters and non-coding RNAs), and they will also be crucial targets for large-scale identification and characterization. It is expected that technologies similar to those described here for enhancers will make it possible to explore their roles in human biology and disease. ■

1. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
2. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
3. Helgadottir, A. *et al.* A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* **316**, 1491–1493 (2007).
4. McPherson, R. *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* **316**, 1488–1491 (2007).
5. Hindorf, L. A., Junkins, H. A., Mehta, J. P. & Manolio, T. A. A catalog of published genome-wide association studies. *OPG: Catalog Published Genome-Wide Assoc. Studies* <<http://www.genome.gov/gwastudies>> (2009).
6. Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional regulatory elements in the human genome. *Annu. Rev. Genomics Hum. Genet.* **7**, 29–59 (2006).
This paper is a comprehensive overview of functional classes of gene regulatory sequence, including many disease-relevant examples identified through gene-centric studies.
7. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a β -globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981).
8. Panne, D. The enhanceosome. *Curr. Opin. Struct. Biol.* **18**, 236–242 (2008).
9. Visel, A., Bristow, J. & Pennacchio, L. A. Enhancer identification through comparative genomics. *Semin. Cell Dev. Biol.* **18**, 140–152 (2007).
10. Visel, A. *et al.* Functional autonomy of distant-acting human enhancers. *Genomics* **93**, 509–513 (2009).
11. Ingram, V. M. Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. *Nature* **180**, 326–328 (1957).
12. Pauling, L. *et al.* Sickle cell anemia, a molecular disease. *Science* **110**, 543–548 (1949).
13. Kan, Y. W. *et al.* Deletion of α -globin genes in haemoglobin-H disease demonstrates multiple α -globin structural loci. *Nature* **255**, 255–256 (1975).
14. Kioussis, D., Vanin, E., deLange, T., Flavell, R. A. & Grosfeld, F. G. β -Globin gene inactivation by DNA translocation in $\gamma\beta$ -thalassaemia. *Nature* **306**, 662–666 (1983).
15. Semenza, G. L. *et al.* The silent carrier allele: β thalassaemia without a mutation in the β -globin gene or its immediate flanking regions. *Cell* **39**, 123–128 (1984).
16. Kleinjan, D. A. & Lettice, L. A. Long-range gene control and genetic disease. *Adv. Genet.* **61**, 339–388 (2008).
17. Sagai, T., Hosoya, M., Mizushima, Y., Tamura, M. & Shiroishi, T. Elimination of a long-range *cis*-regulatory module causes complete loss of limb-specific *Shh* expression and truncation of the mouse limb. *Development* **132**, 797–803 (2005).
This paper shows that deletion of the distant-acting limb enhancer of the *Shh* gene in mice causes severe limb truncation, providing a model example of the requirement for enhancers in mammalian development.
18. Lettice, L. A. *et al.* A long-range *Shh* enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
19. Clark, R. M., Marker, P. C. & Kingsley, D. M. A novel candidate gene for mouse and human preaxial polydactyly with altered expression in limbs of *Hemimelic extra-toes* mutant mice. *Genomics* **67**, 19–27 (2000).
20. Furniss, D. *et al.* A variant in the sonic hedgehog regulatory sequence (ZRS) is associated with triphalangeal thumb and deregulates expression in the developing limb. *Hum. Mol. Genet.* **17**, 2417–2423 (2008).
21. Masuya, H. *et al.* A series of ENU-induced single-base substitutions in a long-range *cis*-element altering Sonic hedgehog expression in the developing mouse limb bud. *Genomics* **89**, 207–214 (2007).
22. Lettice, L. A., Hill, A. E., Devenney, P. S. & Hill, R. E. Point mutations in a distant sonic hedgehog *cis*-regulator generate a variable regulatory output responsible for preaxial polydactyly. *Hum. Mol. Genet.* **17**, 978–985 (2008).
23. Lettice, L. A. *et al.* Disruption of a long-range *cis*-acting regulator for *Shh* causes preaxial polydactyly. *Proc. Natl Acad. Sci. USA* **99**, 7548–7553 (2002).
24. Bolk, S. *et al.* A human model for multigenic inheritance: phenotypic expression in Hirschsprung disease requires both the *RET* gene and a new 9q31 locus. *Proc. Natl Acad. Sci. USA* **97**, 268–273 (2000).
25. Gabriel, S. B. *et al.* Segregation at three loci explains familial and population risk in Hirschsprung disease. *Nature Genet.* **31**, 89–93 (2002).
26. Emison, E. S. *et al.* A common sex-dependent mutation in a *RET* enhancer underlies Hirschsprung disease risk. *Nature* **434**, 857–863 (2005).
27. Grice, E. A., Rochelle, E. S., Green, E. D., Chakravarti, A. & McCallion, A. S. Evaluation of the *RET* regulatory landscape reveals the biological relevance of a HSCR-implicated enhancer. *Hum. Mol. Genet.* **14**, 3837–3845 (2005).
28. Cookson, W., Liang, L., Abecasis, G., Moffatt, M. & Lathrop, M. Mapping complex disease traits with global gene expression. *Nature Rev. Genet.* **10**, 184–194 (2009).
29. Aparicio, S. *et al.* Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes*. *Proc. Natl Acad. Sci. USA* **92**, 1684–1688 (1995).
30. Loots, G. G. *et al.* Identification of a coordinate regulator of interleukins 4, 13 and 5 by cross-species sequence comparisons. *Science* **288**, 136–140 (2000).
31. Nobrega, M. A., Ovcharenko, I., Afzal, V. & Rubin, E. M. Scanning human gene deserts for long-range enhancers. *Science* **302**, 413 (2003).
32. Pennacchio, L. A. *et al.* In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502 (2006).
33. Visel, A. *et al.* Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nature Genet.* **40**, 158–160 (2008).
34. Woolfe, A. *et al.* Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3**, e7 (2005).
35. Bejerano, G. *et al.* Ultraconserved elements in the human genome. *Science* **304**, 1321–1325 (2004).
36. Prabhakar, S. *et al.* Close sequence comparisons are sufficient to identify human *cis*-regulatory elements. *Genome Res.* **16**, 855–863 (2006).
37. Cooper, G. M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **15**, 901–913 (2005).
38. Ahituv, N. *et al.* Deletion of ultraconserved elements yields viable mice. *PLoS Biol.* **5**, e234 (2007).
This paper shows that deletion of several ultraconserved non-coding sequences in mice may not result in obvious phenotypes, demonstrating that even extreme evolutionary constraint does not necessarily indicate that a non-coding sequence is required for viability.
39. The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
40. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genet.* **39**, 311–318 (2007).
This paper identifies a histone H3K4 differential methylation signature that distinguishes promoters from enhancers, providing a chromatin-based tool for genome-wide enhancer prediction.
41. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
42. Xi, H. *et al.* Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet.* **3**, e136 (2007).
43. Wei, C. L. *et al.* A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**, 207–219 (2006).
This paper describes mapping of protein–DNA interactions by ChIP coupled with conventional capillary-based sequencing of concatenated paired-end tags (ChIP–PET), a conceptual predecessor of the ChIP–seq approach.
44. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
45. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of *in vivo* protein–DNA interactions. *Science* **316**, 1497–1502 (2007).

46. Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature Methods* **4**, 651–657 (2007).
This paper is one of several independently published early ChIP-seq studies validating the method for genome-wide mapping of transcription-factor-binding sites.
47. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
This paper is one of several independently published early ChIP-seq studies providing some of the first genome-wide data sets of several histone modifications in different mouse cell types and examining their correlation with functional genome features.
48. Zhao, X. D. *et al.* Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* **1**, 286–298 (2007).
49. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
50. Wederell, E. D. *et al.* Global analysis of *in vivo* Foxa2-binding sites in mouse adult liver using massively parallel sequencing. *Nucleic Acids Res.* **36**, 4549–4564 (2008).
51. Robertson, A. G. *et al.* Genome-wide relationship between histone H3 lysine 4 mono- and tri-methylation and transcription factor binding. *Genome Res.* **18**, 1906–1917 (2008).
52. Ku, M. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* **4**, e1000242 (2008).
53. Cuddapah, S. *et al.* Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* **19**, 24–32 (2009).
54. Gao, N. *et al.* Dynamic regulation of *Pdx1* enhancers by Foxa1 and Foxa2 is essential for pancreas development. *Genes Dev.* **22**, 3435–3448 (2008).
55. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genet.* **40**, 897–903 (2008).
56. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
57. Cooper, G. M. & Brown, C. D. Qualifying the relationship between sequence conservation and molecular function. *Genome Res.* **18**, 201–205 (2008).
58. McGaughey, D. M. *et al.* Metrics of sequence constraint overlook regulatory sequences in an exhaustive analysis at *phox2b*. *Genome Res.* **18**, 252–260 (2008).
59. Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169–181 (2005).
60. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Rev. Genet.* **10**, 57–63 (2009).
61. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
62. Amano, T. *et al.* Chromosomal dynamics at the *Shh* locus: limb bud-specific differential regulation of competence and active transcription. *Dev. Cell* **16**, 47–57 (2009).
63. Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F. & Fraser, P. Long-range chromatin regulatory interactions *in vivo*. *Nature Genet.* **32**, 623–626 (2002).
64. Fullwood, M. J., Wei, C. L., Liu, E. T. & Ruan, Y. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res.* **19**, 521–532 (2009).
65. Miele, A. & Dekker, J. Long-range chromosomal interactions and gene regulation. *Mol. Biosyst.* **4**, 1046–1057 (2008).
66. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genet.* **38**, 1348–1354 (2006).
67. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genet.* **38**, 1341–1347 (2006).
68. Dostie, J. *et al.* Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16**, 1299–1309 (2006).
69. Lee, T. I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301–313 (2006).
70. Squazzo, S. L. *et al.* Suz12 binds to silenced regions of the genome in a cell-type-specific manner. *Genome Res.* **16**, 890–900 (2006).
71. Van Lente, F., Jackson, J. F. & Weintraub, H. Identification of specific crosslinked histones after treatment of chromatin with formaldehyde. *Cell* **5**, 45–50 (1975).
72. Solomon, M. J., Larsen, P. L. & Varshavsky, A. Mapping protein–DNA interactions *in vivo* with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell* **53**, 937–947 (1988).
73. Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309 (2000).
74. Iyer, V. R. *et al.* Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**, 533–538 (2001).
75. Horak, C. E. *et al.* GATA-1 binding sites mapped in the β -globin locus by using mammalian chip-chip analysis. *Proc. Natl Acad. Sci. USA* **99**, 2924–2929 (2002).
76. Barski, A. & Zhao, K. Genomic location analysis by ChIP-seq. *J. Cell. Biochem.* **107**, 11–18 (2009).

Acknowledgements We thank M. Blow, S. Deutsch and A. Sczyrba for help with computational analysis of GWAS data and C. Attanasio for comments. L.A.P. and E.M.R. were supported by the Berkeley Program for Genomic Applications (funded by the US National Heart, Lung, and Blood Institute), and the Director, Office of Science, Office of Basic Energy Sciences, US Department of Energy, under contract number DE-AC02-05CH11231. L.A.P. was also supported by the US National Human Genome Research Institute.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence should be addressed to L.A.P. (lapennacchio@lbl.gov).

Implications of chimaeric non-co-linear transcripts

Thomas R. Gingeras¹

Deep sequencing of 'transcriptomes' — the collection of all RNA transcripts produced at a given time — from worms to humans reveals that some transcripts are composed of sequence segments that are not co-linear, with pieces of sequence coming from distant regions of DNA, even different chromosomes. Some of these 'chimaeric' transcripts are formed by genetic rearrangements, but others arise during post-transcriptional events. The 'trans-splicing' process in lower eukaryotes is well understood, but events in higher eukaryotes are not. The existence of such chimaeric RNAs has far-reaching implications for the potential information content of genomes and the way it is arranged.

Recent years have seen an increasing appreciation of the pervasive (genome-wide) transcription that occurs in most genomes and the multiple functional roles of RNAs within cells^{1,2}. Many individual laboratories have contributed to this work, although much of it was done for large projects, such as the Encyclopedia of DNA Elements (ENCODE)³, the model genome ENCODE (modENCODE)⁴, FANTOM⁵ and the US National Institutes of Health's Roadmap Epigenome⁶. One of the goals of these consortia was to map and characterize comprehensively the transcriptomes of humans and a diverse collection of model organisms. The relatively rapid development of powerful high-throughput sequencing⁷ and *in vivo* screening techniques⁸ has provided the means to accomplish these goals in a very short time. Such studies have consistently found that non-protein-coding RNAs make up most of the transcriptional output of genomes and that many of the functional systems operating within cells involve non-coding transcripts^{9,10}. Additionally, mapping and sequence characterization of the RNAs involved in many of these cellular processes have resulted in interesting observations about the way the information stored in genomes is organized, the way the expression of these RNAs is regulated, novel biochemical processes that point to a possible function and fate of new RNA classes, and the evolutionary implications of increased complexity in the way the genome is organized. In turn, the involvement of RNAs in each of these areas has prompted a reconsideration of the roles of RNA in inheritance and disease^{11,12}. One striking class of non-coding RNA found in several species consists of transcripts containing either sequence elements that map at large distances from each other or that are found on separate chromosomes. Such chimaeric RNAs reveal new dimensions to how information can be stored in genetic systems. This Review surveys the regulatory and evolutionary implications of such chimaeric RNAs.

Organization of information in genomes

In eukaryotic cells, the storage units of heritable information are compartmentalized in the nucleus as DNA and as epigenetic marks etched on the DNA itself and the associated chromatin proteins. RNA, by contrast, has been seen primarily as an intermediary that accurately transfers information from the genome. Its functional roles, such as assisting in the synthesis of proteins by acting as a messenger (mRNA) or as a scaffold for protein synthesis (rRNAs), and helping to gather amino acids for protein synthesis (tRNA), have encouraged the view

that the information stored in the genome is transferred to RNA in a co-linear fashion; in other words, the nucleotide sequences found in RNA transcripts are ordered in the same linear fashion as those found in the DNA genome. The discovery of splicing provided a more modular and non-contiguous view of this co-linear relationship, but even so the order of sequences in both DNA and RNA has been maintained (Fig. 1a). This co-linear organization seems logical and efficient given the perceived primacy of DNA in the genetic hierarchy. Additionally, underlying this co-linear organization of information in the genome and its transfer to RNA is the premise that the sequences that will be joined together in the mature RNAs reside on the same precursor RNA molecule. Indeed, this seems to be the primary path of RNA processing from primary to mature transcripts.

However, structural studies of RNAs in several species have revealed that the sequences that are ultimately joined together on the same mature transcript can be encoded in separately transcribed RNAs with multiple distinct genomic origins (Fig. 1b). Individual RNAs can be transcribed on separate chromosomes (form 1 in Fig. 1b), on the same chromosome but with a different genomic order from that found in the mature RNA (form 2), on the same chromosome but transcribed from different strands (form 3), or on the same chromosome but from different alleles (form 4).

One of the first observations that supported the joining of sequences derived from separate molecules was the discovery of *trans*-splicing. Discovered first in trypanosomes¹³, the process was later shown to occur in *Caenorhabditis elegans* and other nematodes¹⁴, in the protist *Euglena*¹⁵, in flatworms^{16,17} and in higher eukaryotes¹⁸. In nematodes and trypanosomes, short stretches of nucleotides mapping to separate distal regions in their genomes are *trans*-spliced onto the 5' ends of many protein-coding genes. These short sequences are derived from short leader RNA genes (SL genes) that can be positioned thousands of nucleotides away, upstream or downstream (5' or 3', respectively), or on different chromosomes from the genes to which their sequences are spliced. Additionally, unlike the *cis*-splicing observed in higher eukaryotic cells, this *trans*-splicing involves the cleavage and joining of two separate transcripts.

Trans-splicing is not limited to worms and trypanosomes. Clear evidence of *trans*-splicing in mammalian cells is beginning to emerge. One of the most striking examples was reported in human cells by Li *et al.*¹⁹, with the characterization of a developmentally regulated *trans*-splicing

¹Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA.

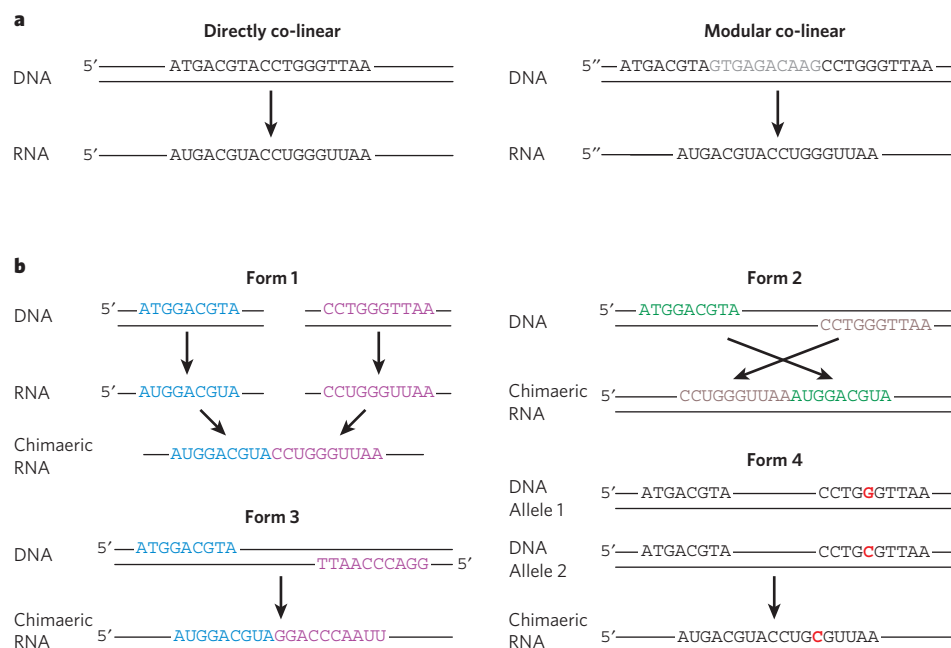


Figure 1 | Models of possible organization of the information in DNA and its transfer to RNA. **a**, Co-linear alignments can be categorized in two forms. In directly co-linear arrangements, information (sequence) is transferred in an uninterrupted fashion to RNA, as is seen in most bacterial mRNA. In modular alignments, information is transferred to RNA in a co-linear fashion but is interrupted (by introns, grey). **b**, There are four possible forms of non-co-linear alignment. The production of precursor RNA is shown only for form 1 but occurs in all forms. Form 1 represents the formation of chimaeric

RNAs from two different loci at two different genic regions in a genome. The precursor RNAs are processed (for example, by *trans*-splicing or RNA recombination¹⁸) into the chimaeric RNA. Form 2 exemplifies the formation of a chimaeric RNA from the same gene by the rearrangement of exons (see Fig. 2 for example). Form 3 illustrates the formation of a chimaeric RNA from two precursor RNAs transcribed from opposite strands. Form 4 describes the formation of a chimaeric RNA from transcripts derived from two alleles of the same gene, one of which contains a single-nucleotide polymorphism (G→C).

event involving the 5' exons of transcripts from the *JAZF1* gene on chromosome 7p15 and the 3' exons of *JJAZ1* (also known as *SUZ12*) located on chromosome 17q11. In endometrial stroma cells, this chimaeric RNA is translated into a chimaeric anti-apoptotic protein. Interestingly, neoplastic stroma cells of the endometrium constitutively express the chimaeric RNA and protein because of a translocation involving these two chromosomal regions. So similar chimaeric RNA and proteins are made in normal cells and genetically rearranged neoplastic cells by two different, non-overlapping mechanisms. The expression level of the chimaeric RNA in the cells containing the translocation is elevated, and this loss of regulation of the expression of the chimaeric gene products is similar to regulatory mutations in oncogenes associated with neoplastic transformation²⁰. These studies also raise the possibility that the *trans*-splicing events may be a precondition for chromosomal exchange and further suggest studies to determine whether chimaeric RNAs, which are often found as a result of such structural mutations, are present in the same non-transformed cells.

A second example of a clinically important *trans*-spliced RNA was observed in a study of a novel erythroid transformation specific (ETS) family fusion transcript *SLC45A3–ELK4* (ref. 21). In this study, the chimaeric *SLC45A3–ELK4* transcript was expressed in normal and benign cancer cells of the prostate. Characterization of the fusion mRNA revealed a major variant in which *SLC45A3* exon 1 is fused to *ELK4* exon 2, leading to the expression of a novel protein. Based on quantitative PCR analyses of DNA, unlike other ETS fusions described in prostate cancer, the expression of *SLC45A3–ELK4* mRNA is not exclusive to the formation of chromosomal rearrangements. As with the *JAZF1* and *JJAZ1* case, both the chimaeric RNA and protein can be formed by either *in vivo trans*-splicing or a genetic rearrangement event. The *SLC45A3–ELK4* chimaeric RNA result is notable for two clinical reasons. First, the chimaeric transcript can be detected at high levels in urine samples from men at risk of prostate cancer. Second, treatment of the LNCaP cell line with R1881 synthetic androgen indicates that the fusion transcript is differentially regulated, making treatment with androgen antagonists of potential therapeutic value.

The increased prevalence of chimaeric RNAs in other normal cells is supported by both clinical and empirical observations. Cancer geneticists have puzzled over why chimaeric RNAs are detected in normal tissues used as controls for transformed cells. As with *JAZF1* and *JJAZ1* genes, chimaeric RNAs involving the immunoglobulin heavy chain gene (*IGH*) and the *BCL2* gene have been observed in spleen tissues from normal individuals. Interestingly, a translocation (t(14;18)) involving the same genes is seen in neoplastic haematopoietic cells such as lymphomas²². Additionally, in other studies designed to identify the 5' termini of genes expressed in many different tissues and cell lines as part of the pilot phase of the ENCODE project, approximately 65% of the genes tested were found to be involved in the formation of chimaeric RNAs. These genes had distal unannotated 5' transcription start sites that were joined to RNAs from genes located hundreds of thousands of nucleotides 3' to these sites. In such cases, multiple intervening genes and their corresponding transcripts were not involved in the formation of the chimaeric RNAs. Although these results could be examples of very-long-distance co-linear splicing, this is unlikely because some of the genes that take part in the formation of the chimaeric RNAs are encoded on the opposite strand. However, unlike the studies done by Li *et al.*¹⁹, these normal tissues were not tested for the presence of low-level rearrangement events. Two studies have recently suggested that detection of chimaeric RNAs by the deep sequencing of individuals' transcriptomes could provide early detection of the kinds of chromosomal rearrangements found in patients with cancer²³. Although the detection of chimaeric RNAs may sometimes result from a genetic translocation event, it is misleading to assume that such rearrangements are present, because the chimaeric RNAs could have come from non-chromosomal rearrangements.

These and other studies raise the question of whether chimaeric RNAs that contain sequences from multiple non-co-linear positions in the genome are more commonly found in normal tissues than previously appreciated. Evidence from the pilot phase of the ENCODE project^{3,24,25} and other independent studies (see, for example, refs 26 and 27) supports the prevalence of such chimaeric RNA production in both normal

tissues and established cell lines¹⁸. The pilot ENCODE studies reported that approximately 65% of the protein-coding genes mapping within the boundaries of the ENCODE circumscribed regions contributed to the formation of chimaeric RNAs, half of which contained sequences of genes that mapped to the opposite strand of the index gene used in these studies (unpublished observations). It is important to note that the observation of chimaeric RNAs involving a large proportion of genes does not imply that such RNAs have high copy numbers. However, the level of expression of exons comprising chimaeric RNAs is often similar to that of annotated transcripts²⁴.

An unusual type of chimaeric transcript has been reported and is represented by chimaeric RNAs that contain exon sequences that were out of order compared with the exons found in the human genome. This type of non-co-linear representation of information in the genome is exemplified by a *SEC14L2* isoform in which upstream exons were inserted between downstream exons (Fig. 2). Spliced portions of exons 3, 4 and 5 were inserted downstream of exons 6, 7 and 8 and connected to the 3' untranslated region (UTR) of the downstream gene *HSPC242* (ref. 28). This arrangement preserves a large open reading frame while putting amino acids 56–93 of *SEC14L2* after amino acid 218 of the protein.

The molecular mechanisms controlling the joining of sequences from two individual RNA molecules are still uncertain, although we do have some leads. For example, Bruzik and Maniatis showed that RNA molecules containing a 3' splice site and enhancer sequence are efficiently spliced in *trans*^{29,30}. The products are RNA molecules that contain normally *cis*-spliced 5' splice sites or are *trans*-spliced, as seen with the SL RNAs from lower eukaryotes. Additionally, Li *et al.*¹⁸ computationally identified chimaeric RNAs in the RNA and expressed sequence tag (EST) databases of yeast, fly, mouse and human, and have confirmed approximately 30% of them by RT-PCR. They frequently found short homologous sequences (SHSs) at the junction sites of the chimaeric RNAs and, curiously, they proposed a transcriptional-slippage model, rather than the 'copy choice'³¹ or classic *trans*-splicing model³², to explain the generation of those chimaeric RNAs synthesized from templates with SHSs.

Whatever mechanism(s) controls the formation of chimaeric RNAs, they are seen in eukaryotic cells from worms to humans. Nevertheless, with the exception of some well-studied cases cited previously, no function has yet been identified for the large majority of observed chimaeric RNAs.

The expression of non-co-linear information

The organization of information in a genome and its non-co-linear transfer into RNA has implications for the way in which such transcriptional processes are regulated. If information transferred to RNA can

be derived from sequences found in at least two different transcripts, then either the regulation of the synthesis of the individual transcripts involved in the *trans*-splicing must be coordinated or the half-life characteristics of at least some transcripts must allow temporally uncoordinated regulation of their expression. The coordinated regulation of expression of individual transcripts and the subsequent joining of these individual transcripts raises the possibility that these activities are carried out in close three-dimensional proximity in the cell.

Studies of nuclear organization in a cell indicate that non-ribosomal RNAs may be transcribed in subcompartments or foci known as transcription factories³³ and that these factories are stably maintained in the absence of active transcription³⁴. Transcription factories are sites enriched in RNA polymerase II and may contain other processing factors needed to create mRNAs^{35,36}. For any single cell, there seem to be fewer transcriptional factories than the number of expressed genes³⁷, which suggests that multiple genes share the same transcriptional machinery. Genes located *cis* and *trans* to one another may share the same factory, suggesting that distal genes migrate to preassembled nuclear sites. Osborne *et al.* used fluorescent *in situ* hybridization to investigate the genes associated with transcription factories during immediate-early gene induction in mouse B lymphocytes³⁸. They found that the mouse *Myc* proto-oncogene, on chromosome 15, is recruited to the same transcription factory as the highly transcribed *Igh* gene, located on chromosome 12. A similar association is seen in human cells³⁸. Interestingly, human *MYC* and *IGH* are the most frequent translocation partners in plasmacytoma and Burkitt's lymphoma. These data suggest that there is a direct link between the non-random association of two genes in an interchromosomal organization of transcribed genes in transcription factories and suggest that this association may be involved in specific chromosomal translocation. Although chimaeric RNAs have not yet been observed in normal plasmacytes, the three-dimensional association of these two genes to the same transcriptional factory offers the possibility that the RNAs made from *MYC* and *IGH* participate in *trans*-splicing, and that there may be chromosomal translocation events, as seen with the *JAZF1* and *JJAZ1* and *SLC45A3* and *ELK4* genes described earlier. Transcription factories that can carry out transcription and the formation of chimaeric RNAs may be either normal or specialized. Evidence for specialized factories has been described³⁷, and the nucleolus is an excellent example. Figure 3 shows a model for a hypothetical specialized factory. Such a model predicts a non-random correlation among the genomic sites that can be observed in close proximity in three-dimensional space, as detected by chromosome conformation capture carbon copy (5C)³⁹ and genomic regions encoding transcripts that are involved in the formation of chimaeric RNAs.

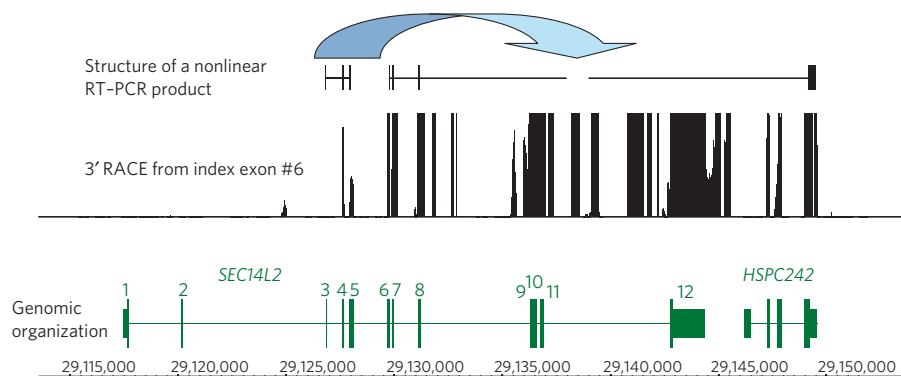


Figure 2 | Characterization of a chimaeric transcript. The organization of the genomic region containing the genes *SEC14L2* and *HSPC242* is shown (bottom), including the exons that have been annotated and a scale showing the nucleotide position on human chromosome 22. The structure of an RT-PCR product corresponding to this region is also shown (top). In the cDNA derived from this product, sequences corresponding to exons 3, 4 and 5 of the *SEC14L2* gene were found to be *trans*-spliced (blue arrow) to the 3' side (downstream) of exon 8 of this gene. The sequence immediately

following exon 5 is the last exon of the *HSPC242* gene. The sequence of this *trans*-spliced cDNA was discovered by first carrying out a 3' RACE (rapid amplification of cloned/cDNA ends) reaction using exon 6 as the index position of the RACE primer. After hybridizing the 3' RACE products, the array results are depicted in the second trace. Note that although exons 4 and 5 map upstream of the index site used in the RACE reaction, they are detected on the array profile. This is because the RACE reaction is copying an RNA containing the *trans*-spliced exons.

Evolutionary implications

Both the information content of genomes and the strategies to transfer this information from DNA to RNA have increased over the course of evolution. The transfer of information in eukaryotic cells from worms to humans has also been accompanied by increased segmentation of genomic information. This segmentation has allowed permutations of sequences in an RNA context and the development of increasingly complex RNA-processing mechanisms, as demonstrated by RNA splicing and *trans*-splicing. The segmentation of information in a DNA context, and the mechanisms to create permutations of them for RNA, are potential areas of active selection. The possible use of multiple precursor transcripts and the ability to alter the exon combinations found within them to form mature chimaeric transcripts would have the advantage of increasing the informational content of genomes dramatically, and would probably drive an increase in the complexity of the processing machinery. For a single primary transcript, this increase in complexity can be expressed by formula 1 in Box 1 and is illustrated in Fig. 4a. However, in the case of two or more primary transcripts contributing to a mature processed RNA in a non-co-linear and permuted fashion, the total number of possible transcripts can be expressed by formula 3 in Box 1 and is illustrated in Fig. 4b, c.

Using DNA as a reference for evolutionary studies

The potential use of multiple precursor transcripts to form a single mature chimaeric RNA product, and the possibility of altering the order of the genomic sequences in chimaeric transcripts, raises two points that have evolutionary implications. The first concerns the identification of a complete set of both functional sequence segments and the evolutionarily constrained sequences. Constrained genomic sequences are often identified by the alignment of orthologous sequences using one of several sequence-analysis programs^{40–45}. Such alignments are often used to infer functionality. However, if functional RNA sequence regions can be made from modules of non-co-linear sequences that arise by the permutation of genomic sequences encoded in the DNA,

short evolutionary constrained or novel functional regions of genomes may be joined together in RNA space. Such products cannot be found by analysing genomic sequences. Programs designed to identify constrained sequences in DNA would probably fail to detect such novel RNA-encoded sequences.

The second point concerns the use of genomic DNA as the sole reference to identify the sequences that manifest the effects of evolutionary pressures. Using DNA as a reference to study evolutionary pressures on sequences naturally stems from its role in inheritance. However, focusing on DNA as the sole subject to catalogue genetically transmitted functional elements, and as the only molecule to analyse for evolutionary constraints, ignores the growing number of acknowledged functional roles of RNA. It also potentially misses the novel RNA-encoded sequence elements that are often used to carry out these functions. The identification of functional regions encoded in a co-linear fashion is relatively straightforward for both DNA and RNA, but the same cannot be said for functional regions created from non-co-linear segments. Unless these sequence regions have been catalogued previously (such as the SL1 and SL2 transcripts in *C. elegans*), or the rules of how distal non-co-linear sequences are to be joined in a transcript are known, then analyses of the transcriptome, as well as the genome, may be required if we are to catalogue all the functional regions subject to selection.

Interpreting genetic variation

Genetic variation, from a single nucleotide to gross chromosomal rearrangements, is the basis of evolutionary change. Two biases influence the interpretation of the effects of such genetic variation. The first is that the effect is most often interpreted to be related to the nearest gene annotation. Thus, a variation in an exon of a gene is often interpreted as causing a malfunction of that gene, leading to an abnormal phenotype. The second bias is that the effect of the variation is associated in the context of the extended but local genetic locus. If a genetic variation is observed to be proximal but is not within an exon of a gene, the effects are usually explained in terms of misregulation

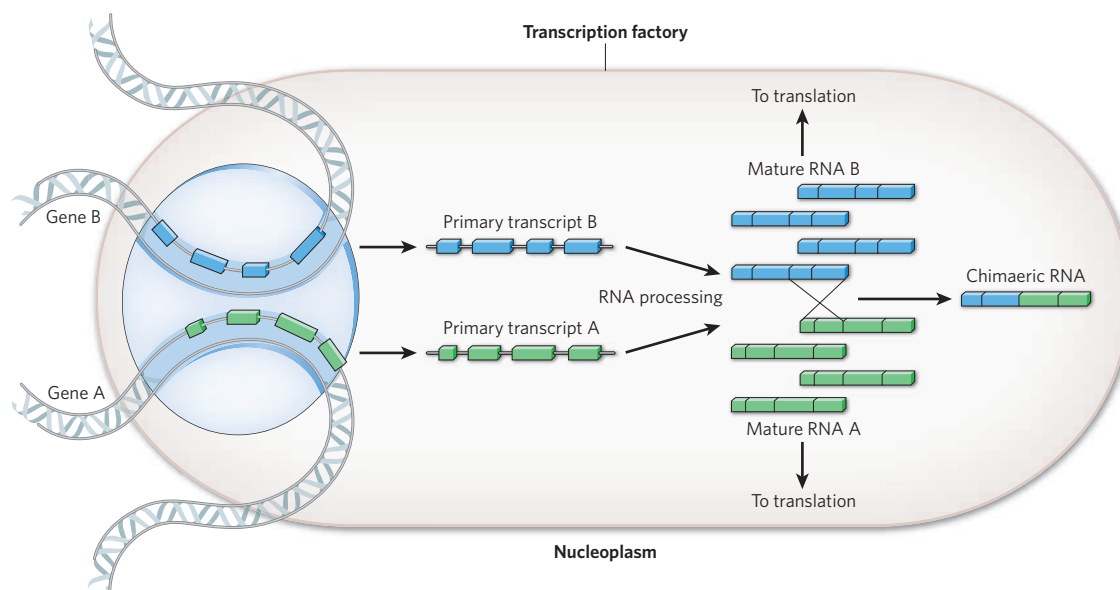


Figure 3 | Model specialized transcription factory for transcription and the formation of chimaeric RNAs. Based on the observation that there are fewer transcription factories in nuclei than the number of transcribed loci, and considering the existence of specialized transcription factories³⁷, this model hypothesizes that genes A and B, encoded in different regions of the genome, are collected in a transcription factory and transcribed into primary transcripts by multiple RNA polymerase II transcriptional complexes (blue oval). The primary transcripts are then processed to form mature spliced and chimaeric RNAs. In this model, most of the primary

RNAs are involved in *cis*-splicing and are transported to the cytosol for translation. Consistent with steady-state estimates of the number of chimaeric RNAs, a small proportion of the primary RNAs are used to create chimaeric RNAs. Either a single isoform or multiple isoforms of chimaeric RNAs can be made from combinations of primary RNA transcribed in the same factory. The occurrence of translocation events involving genomic regions that are transcribed to produce chimaeric RNAs raises the possibility that such rearrangements may also be made at these factories.

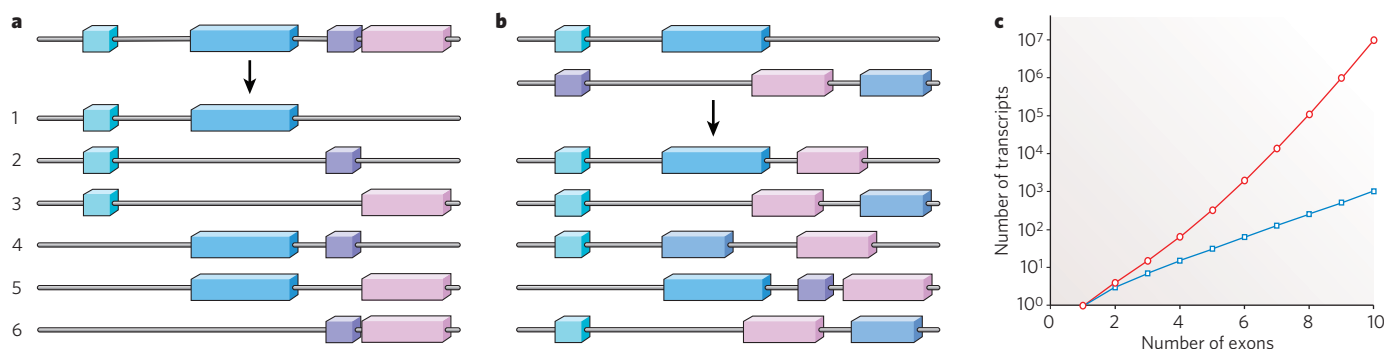


Figure 4 | Co-linear and non-co-linear combinations of modules of information. **a**, All six possible two-exon combinations derived from a four-exon genic region are shown. **b**, There are 325 possible non-co-linear, permuted three-exon transcripts that can be made from two transcripts containing two and three exons; only five of them are shown here. **c**, A semi-log

plot comparing the number of transcript isoforms that can be created using 1–10 exons for a non-permuted, co-linear organization of information (blue) and a permuted, non-co-linear organization (red). The total number of mature transcripts for the non-co-linear permuted case diverges exponentially from the co-linear case as the number of exons increases.

of a gene located nearby. Such interpretations are common and are often well supported. However, current genetic association studies are widening the range of study to include long-distance trans effects, for example with single-nucleotide polymorphisms mapping hundreds of thousands of base pairs associated with urinary bladder and prostate cancers^{46,47}. This is prompted by the recent observation that almost half of all polymorphisms found to be statistically correlated with complex diseases or traits are located at a non-annotated and distal genomic site (an intron or intergenic region)⁴⁸. The joining of distal genomic regions by the formation of chimaeric RNAs provides additional motivation for this broadening of genetic analyses.

Conclusions

Consideration of the non-co-linear organization of genomes raises several questions. For example, why do genomes resort to a non-co-linear strategy of organizing stored information? The relatively straightforward co-linear organization of functional information in the genome

is likely to be most common, but non-co-linear organization appeared early in the evolution of the eukaryotic genome, as exemplified by worms and trypanosomes. The advantages of such a complex strategy include providing a means of increasing the information content of genomes and allowing possible new combinations of exons operating in a relatively redundant fashion (motif sharing) and functioning in more significant roles (such as the formation of newly spliced mRNAs allowing increased protein diversity). Additionally, the use of non-co-linear organization leading to the formation of chimaeric RNAs might also allow the real-time monitoring of transcripts that are co-regulated RNAs. Such a function leads to the prediction, which remains to be tested, that chimaeric RNAs would be formed from genes that had similar expression profiles observed over development or in response to some external stimuli.

A second question concerns the relatively low-level expression of chimaeric transcripts, as illustrated by the infrequent reports from the many cDNA library studies. Is this because cells rarely use chimaeric RNAs or because such RNAs are non-functional units of transcription and part of a general RNA-processing background? The answer is not yet clear. The lack of multiple reports of biologically functional chimaeric RNAs in organisms other than worms, trypanosomes and a few higher eukaryotes may stem from the fact they have often been observed in cDNA studies but discarded because of the current models of genome organization. Alternatively, perhaps they are simply not observed because of the shallow depth of sampling of the transcriptome or as a result of the method of analysis (such as the use of arrays).

What is clear, however, is that the non-co-linear organization of information in genomes is more common than previously thought. It is seen both in lower eukaryotes, in which *trans*-splicing is integral, and in higher eukaryotes, in which its presence is observed but its importance remains unclear. Its strategic and evolutionary advantages are interesting, but many questions remain. Discussions about the functional importance of the transcription of non-protein-coding DNA are already under way. If we are to understand the nature of the non-co-linear organization of information, additional experiments will be required, and these are likely to entail an expansion of efforts in two underdeveloped areas of study. The first involves the three-dimensional characterization of important biochemical processes in which the spatial proximities of component molecules allow distal regions of the genome or RNAs to interact. The second area involves studying in greater detail subcellular compartmentalization in which specialized molecular processes occur. These subcompartment studies are likely to find both protein and RNA components of specialized processes that are difficult to detect when whole cells are examined (because of their low copy number) but are enriched when subcellular fractionation is carried out. Both of these areas of experimentation will increase our appreciation of the elegance and complexity of genetic systems.

Box 1 | Potential diversity from non-co-linear transcripts

The total number of possible co-linear combinations can be expressed as

$$N(L, m) = \frac{L!}{(L-m)!m!} \quad (1)$$

where L is the total number of possible exons in the primary transcript and m is the number of exons found in the mature transcript. Thus, from a single primary transcript containing 4 possible exons ($L = 4$), the number of mature transcripts that have 2 exons is 6 (Fig. 4a). There are 15 possible co-linear transcripts containing 1–4 exons, or more generally:

$$N(L) = \sum_{m=1}^L N(L, m) = 2^L - 1 \quad (2)$$

If two or more primary transcripts contribute to a mature processed RNA in a non-co-linear and permuted fashion, the total number of possible permuted transcripts can be expressed as

$$N(L, m) = \frac{L!}{(L-m)!} \quad (3)$$

where L again is the total number of exons and m is the number of exons in the mature transcripts. Thus, if two primary transcripts with 2 and 3 exons ($L = 5$) are used to create only a 3-exon mature transcript ($m = 3$), a total of 60 permuted transcripts can be made (Fig. 4b). There are 325 possible permuted transcripts using 1–5 exons. The dependence of the total number of mature transcripts on the number of possible exons for the non-co-linear permuted case diverges exponentially from the co-linear case (Fig. 4c).

1. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nature Rev. Genet.* **10**, 155–159 (2009).
2. Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **136**, 629–641 (2009).
3. ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
4. Celniker, S. E. *et al.* Unlocking the secrets of the genome. *Nature* **459**, 927–930 (2009).
5. Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).
6. Jones, P. A. *et al.* Moving AHEAD with an international human epigenome project. *Nature* **454**, 711–715 (2008).
7. Hert, D. G., Fredlake, C. P. & Barron, A. E. Advantages and limitations of next-generation sequencing technologies: a comparison of electrophoresis and non-electrophoresis methods. *Electrophoresis* **29**, 4618–4626 (2008).
8. Ramadan, N., Flockhart, I., Booker, M., Perrimon, N. & Mathey-Prevot, B. Design and implementation of high-throughput RNAi screens in cultured *Drosophila* cells. *Nature Protocols* **2**, 2245–2264 (2007).
9. Amaral, P. P. & Mattick, J. S. Noncoding RNA in development. *Mamm. Genome* **19**, 454–492 (2008).
10. Mattick, J. S., Amaral, P. P., Dinger, M. E., Mercer, T. R. & Mehler, M. F. RNA regulation of epigenetic processes. *Bioessays* **31**, 51–59 (2009).
11. Cooper, T. A., Wan, L. & Dreyfuss, G. RNA and disease. *Cell* **136**, 777–793 (2009).
12. Ashe, A. & Whitelaw, E. Another role for RNA: a messenger across generations. *Trends Genet.* **23**, 8–10 (2007).
13. Sutton, R. E. & Boothroyd, J. C. Evidence for *trans* splicing in trypanosomes. *Cell* **47**, 527–535 (1986).
14. Krause, M. & Hirsh, D. A *trans*-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* **49**, 753–761 (1987).
15. Tessier, L. H. *et al.* Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by *trans*-splicing in *Euglena*. *EMBO J.* **10**, 2621–2625 (1991).
16. Rajkovic, A., Davis, R. E., Simonsen, J. N. & Rottman, F. M. A spliced leader is present on a subset of mRNAs from the human parasite *Schistosoma mansoni*. *Proc. Natl Acad. Sci. USA* **87**, 8879–8883 (1990).
17. Davis, R. E. *et al.* RNA *trans*-splicing in *Fasciola hepatica*. Identification of a spliced leader (SL) RNA and SL sequences on mRNAs. *J. Biol. Chem.* **269**, 20026–20030 (1994).
18. Li, X., Zhao, L., Jiang, H. & Wang, W. Short homologous sequences are strongly associated with the generation of chimeric RNAs in eukaryotes. *J. Mol. Evol.* **68**, 56–65 (2009).
This paper includes computational analyses of human, mouse and fly RNA databases to identify chimaeric RNAs and short repeat sequences located at the junction sites forming the chimaeric transcripts.
19. Li, H., Wang, J., Mor, G. & Sklar, J. A neoplastic gene fusion mimics *trans*-splicing of RNAs in normal human cells. *Science* **321**, 1357–1361 (2008).
This paper identifies chimaeric RNA and protein from *JAZF1* and *JJAZ1* genes in normal and transformed endometrial cells.
20. Eyche, A., Rocques, N. & Pouppnot, C. A new MAFia in cancer. *Nature Rev. Cancer* **8**, 683–693 (2008).
21. Rickman, D. S. *et al.* *SLC45A3-ELK4* is a novel and frequent erythroblast transformation-specific fusion transcript in prostate cancer. *Cancer Res.* **69**, 2734–2738 (2009).
This paper identifies chimaeric RNA from the *SLC45A3* and *ELK4* genes in normal and transformed prostate cells.
22. Janz, S., Potter, M. & Rabkin, C. S. Lymphoma- and leukemia-associated chromosomal translocations in healthy individuals. *Genes Chromosomes Cancer* **36**, 211–223 (2003).
23. Maher, C. A. *et al.* Transcriptome sequencing to detect gene fusions in cancer. *Nature* **458**, 97–101 (2009).
24. Denoeud, F. *et al.* Prominent use of distal 5' transcription start sites and discovery of a large number of additional exons in ENCODE regions. *Genome Res.* **17**, 746–759 (2007).
This paper identifies and characterizes chimaeric RNAs in normal tissue and transformed cell lines mapping in the 1% of the human genome analysed by the ENCODE studies.
25. Djebali, S. *et al.* Efficient targeted transcript discovery via array-based normalization of RACE libraries. *Nature Methods* **5**, 629–635 (2008).
26. Akiva, P. *et al.* Transcription-mediated gene fusion in the human genome. *Genome Res.* **16**, 30–36 (2006).
27. Parra, G. *et al.* Tandem chimerism as a means to increase protein complexity in the human genome. *Genome Res.* **16**, 37–44 (2006).
28. Kapranov, P. *et al.* Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res.* **15**, 987–997 (2005).
29. Buzik, J. P. & Maniatis, T. Spliced leader RNAs from lower eukaryotes are *trans*-spliced in mammalian cells. *Nature* **360**, 692–695 (1992).
30. Buzik, J. P. & Maniatis, T. Enhancer-dependent interaction between 5' and 3' splice sites in *trans*. *Proc. Natl Acad. Sci. USA* **92**, 7056–7059 (1995).
31. Coffin, J. M. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. *J. Gen. Virol.* **42**, 1–26 (1979).
32. Hastings, K. E. SL *trans*-splicing: easy come or easy go? *Trends Genet.* **21**, 240–247 (2005).
33. Jackson, D. A., Hassan, A. B., Errington, R. J. & Cook, P. R. Visualization of focal sites of transcription within human nuclei. *EMBO J.* **12**, 1059–1065 (1993).
34. Mitchell, J. A. & Fraser, P. Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev.* **22**, 20–25 (2008).
35. Pombo, A. & Cook, P. R. The localization of sites containing nascent RNA and splicing factors. *Exp. Cell Res.* **229**, 201–203 (1996).
36. Pombo, A. *et al.* Regional and temporal specialization in the nucleus: a transcriptionally-active nuclear domain rich in PTF, Oct1 and PIKA antigens associates with specific chromosomes early in the cell cycle. *EMBO J.* **17**, 1768–1778 (1998).
37. Carter, D. R., Eski, C. & Cook, P. R. Transcription factories. *Biochem. Soc. Trans.* **36**, 585–589 (2008).
38. Osborne, C. S. *et al.* Myc dynamically and preferentially relocates to a transcription factory occupied by *Igh*. *PLoS Biol.* **5**, e192 (2007).
39. Dostie, J., Zhan, Y. & Dekker, J. Chromosome conformation capture carbon copy technology. *Curr. Protocols Mol. Biol.* Ch. 21, 14 (2007).
40. Margulies, E. H., Program, N. C. S. & Green, E. D. Detecting highly conserved regions of the human genome by multispecies sequence comparisons. *Cold Spring Harb. Symp. Quant. Biol.* **68**, 255–263 (2003).
41. Margulies, E. H. *et al.* An initial strategy for the systematic identification of functional elements in the human genome by low-redundancy comparative sequencing. *Proc. Natl Acad. Sci. USA* **102**, 4795–4800 (2005).
42. Margulies, E. H. *et al.* Analyses of deep mammalian sequence alignments and constraint predictions for 1% of the human genome. *Genome Res.* **17**, 760–774 (2007).
43. Brudno, M. *et al.* LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* **13**, 721–731 (2003).
44. Cooper, G. M. *et al.* Characterization of evolutionary rates and constraints in three mammalian genomes. *Genome Res.* **14**, 539–548 (2004).
45. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
46. Kiemeny, L. A. *et al.* Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nature Genet.* **40**, 1307–1312 (2008).
47. Yeager, M. *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nature Genet.* **39**, 645–649 (2007).
48. Altshuler, D., Daly, M. J. & Lander, E. S. Genetic mapping in human disease. *Science* **322**, 881–888 (2008).

Acknowledgements Work in my laboratory is supported by the National Human Genome Research Institute (grants U54 HG004557 and U01 HG004271). I thank P. Kapranov for discussions and long-term collaboration, H. Sussman for helpful discussions and editing of the manuscript, and A. Dobin for help deriving the computational expressions to determine the co-linear and non-co-linear permutations, and for the plot shown in Fig. 4c.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence should be addressed to T.R.G. (gingeras@cshl.edu).

Chromosome crosstalk in three dimensions

Anita Göndör¹ & Rolf Ohlsson¹

The genome forms extensive and dynamic physical interactions with itself in the form of chromosome loops and bridges, thus exploring the three-dimensional space of the nucleus. It is now possible to examine these interactions at the molecular level, and we have gained glimpses of their functional implications. Chromosomal interactions can contribute to the silencing and activation of genes within the three-dimensional context of the nuclear architecture. Technical advances in detecting these interactions contribute to our understanding of the functional organization of the genome, as well as its adaptive plasticity in response to environmental changes during development and disease.

The complexity of chromosome architecture has been known about since the end of the nineteenth century, when chromatin loops were first visualized¹. Subsequently, the early *Drosophila* geneticists perceived the importance of interactions along the chromatin fibre, when they observed that inactivation of genes could spread over huge distances along the chromosome in *cis*, and variably from cell to cell, to give rise to variegated gene-silencing effects. These geneticists also had a highly developed understanding of the *trans* effects of chromatin interactions². For example, in 1954, the term 'transvection' was applied to the complementation seen when two alleles of the bithorax complex were paired. When the alleles were separated, this complementation was lost, thereby providing evidence that one allele had to somehow sense its partner allele in order to be active².

In subsequent decades, exploring communication among chromatin fibres remained largely outside the mainstream of chromatin research, which focused on understanding the structure of the chromatin fibre itself. This work uncovered important features, such as the smallest chromatin unit (the nucleosome) and how the primary chromatin fibre is organized into nucleosome arrays. However, the pioneering work by *Drosophila* geneticists eventually led to efforts to explore higher-order chromatin organization within the architecture of the nucleus. Today, it is clear that highly sophisticated but poorly understood processes organize higher-order chromatin structures. These structures in turn contribute to the regulation of transcriptional programs, as well as replication patterns, in the context of the three-dimensional space of the nucleus³.

The nucleus displays an immensely complex architecture that in many instances can be visualized only by using specific antibodies⁴. With the exception of the nuclear lamina, there are no membranes surrounding subcompartments, such as the nucleolus. These structural and functional hallmarks, probably organized stochastically by self-assembling factors⁵, provide key environments for chromatin interactions, as exemplified by active ribosomal RNA gene clusters driving the formation of the nucleoli⁶. Furthermore, large heterochromatic regions of chromosomes assemble at the nuclear lamina in a cell-type-specific manner^{7,8}, whereas transcriptionally active regions tend to loop out into the interior of the nucleus^{7,8}. The simultaneous juxtaposition of active transcriptional units in transcription factories^{9–12} and replicons in replication factories^{13–15} provides yet more levels of organization.

Recent progress in this research area has been facilitated by the development of new methods that allow genome-wide screens of chromosomal interactions. Here, with a focus on mammalian cells, we discuss the current understanding of how chromatin communicates with itself — which we

term chromatin crosstalk — and how this functionally relates to biological processes. We start with an overview of important principles governing chromatin crosstalk in *cis* (loops) and in *trans* (bridges).

Constraints on chromatin crosstalk

Chromatin loops bring distal elements of the chromosome into close physical proximity, with potential consequences for gene expression and/or propagation of the genome. The loops can be visualized when two or more portions of a chromatin fibre interact in *cis*. To enable loop formation, the chromatin fibres must physically encounter each other. A growing body of evidence suggests that stochastic movements of chromatin fibres provide such opportunities by bringing physical neighbours together¹⁶, with the frequency of interactions largely dictated by their proximities to, and affinities for, each other. The physical interactions of chromatin fibres can be measured by using techniques based on chromosome conformation capture (3C) and the related techniques circular chromosome conformation capture (4C) and chromosome conformation capture carbon copy (5C) (Box 1). Analyses of the regulatory regions of the *H19* and β -globin gene loci by using 4C and 5C^{17,18} have revealed large domains of interacting chromatin fibres. Depending on the resolution of the technique, domains encompassing between 100 kilobases (kb) and more than 10⁶ kb have been observed to be in physical proximity.

Shorter-range interactions are restricted by the physical properties of chromatin, with a minimal estimated length of 10 kb for uninterrupted chromatin fibres and 0.5 kb for naked DNA¹⁸. By extrapolation, nucleosome-free regions at promoters and enhancers, for example, provide potential 'hinges', which could increase the mobility of the flanking chromatin fibres, thereby facilitating the formation of shorter-range chromatin loops (Fig. 1). Additionally, as histone-acetylation states regulate chromatin flexibility¹⁹, the formation of chromatin loops may be facilitated by transcription factors cooperating with chromatin-remodelling complexes. Associated helicases, such as CHD8, might help to release torsional stress that hampers the formation of, or results from, chromatin fibre interactions²⁰. For longer-range interactions in *cis*, the chromosome might transiently display particular conformations, facilitated by increased mobility, that bring distal elements into sufficient proximity to promote direct interaction. Patterns of intrachromosomal chromatin folding may therefore be influenced by the position the chromosome occupies in the nucleus and, hence, its neighbourhood.

Whereas chromatin loops describe short-range and long-range interactions in *cis*, chromatin bridges depict long-range interactions in

¹Department of Microbiology, Tumor and Cell Biology, Nobels väg 16, Box 280, Karolinska Institute, SE-171 77 Stockholm, Sweden.

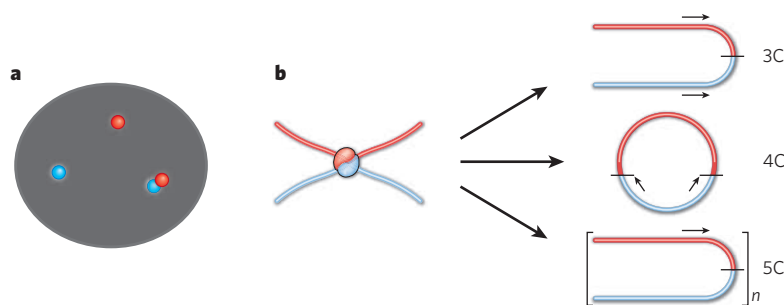
Box 1 | Chromosome conformation capture methods

The 3C method was invented to address the folding of chromosomes and how the chromatin fibre can establish both intrachromosomal and interchromosomal interactions³⁸. Its resolution is higher than that of DNA fluorescence *in situ* hybridization (FISH) analysis by two orders of magnitude but, in contrast to DNA FISH, it does not provide a quantitative assessment of frequencies of physical juxtapositions. Colour-coded DNA FISH analysis visualizing the proximity between alleles of two different loci is shown in panel **a** of the figure (one locus in red, the other in blue). The 3C method has proved useful for determining the close physical proximity of sequences (with a resolution of a few kilobase pairs) from remote interchromosomal or intrachromosomal locations³⁸. Briefly, formaldehyde-crosslinked chromatin is solubilized by detergents, digested with restriction enzymes of choice, and then ligated under very dilute conditions, which favour intramolecular ligation events. Subsequently, interacting chromatin fibres can be identified on reverse-crosslinked ligated DNA by using PCR primers representing both sequences (see figure, panel **b**; small arrows depict PCR primers). The 3C method is, however, less suitable for screening interactions without prior knowledge or expectation of their existence.

To deal with this shortcoming, several laboratories developed alternative methods based on the 3C approach; these are

collectively known as 4C methods. These techniques differ from the 3C method in various ways, such as the inclusion of a circularization step that allows the identification of interacting sequences by using primers positioned on the bait (that is, the known sequence of interest) but close to the junction between the bait and interacting sequence^{25,47} (see figure, panel **b**). This allows high-throughput screening of physical interactions between chromosomes without a preconceived idea of the interacting partners. An interesting variant of 4C methods, termed 5C, is based on analysis of all potential interactions within a limited region and is basically an extended 3C approach (see figure, panel **b**; *n* denotes numerous interactions)¹⁷. However, the 3C, 4C and 5C methods allow at best a semiquantitative estimate of genome-wide patterns of interactions from

particular baits. Moreover, their inability to assess readily the frequency of patterns of interactions necessitates complementing the 4C screens with DNA FISH analysis. The present low sensitivity of the 4C technology (which usually requires at least a million cells) offers only snapshots of accumulated interactions. Thus, it should be expected that the number of interacting elements is limited at any given time and that dynamic on-off patterns of interactions generate a wide range of interacting elements in large cell populations. In this regard, rather than scoring all possible chromatin interactions in the entire genome, which presents both financial and logistical problems, a more prudent strategy is to filter the information with respect to a particular chromatin factor by combining a 3C, 4C or 5C technique with chromatin immunoprecipitation.



trans. One or both partners of chromatin bridges must reach beyond the confines of its chromosome territory (that is, the space occupied by a chromosome in an interphase nucleus) for interactions to be possible²¹. (The potential scenarios for both short-range and long-range interactions are summarized in Fig. 2.) For example, interactions dependent on ligand-activated oestrogen receptors occur at the edges of the relevant chromosome territories. Moreover, the occurrence of such interactions is frequently accompanied by the reorganization of the chromosome territory neighbourhood and depends on β -actin polymerization²².

These observations raise an important unresolved issue, namely that of how, from numerous potential combinations of interactions, it is possible for a locus to select another locus situated on a separate chromosome to interact with specifically. We propose that for precision to be achieved in the interaction, the process occurs in several steps that gradually increase the specificity of the communication between chromatin fibres. The initial step may depend on more general features of larger domains of chromatin, perhaps involving the whole chromosome territory to establish an interaction that is sufficiently stable to promote additional and more specific interactions within the formed complex. Although evidence in support of this idea is scant, chromatin features at individual repeat elements may synergize to create particular constellations of higher-order chromatin conformations and provide a three-dimensional platform for interacting chromatin fibres. In line with this hypothesis, it has been observed that the interchromosomal complex impinging on the interferon- β gene has Alu elements as a common feature²³. Furthermore, several imprinted domains, which can be predicted from specific constellations of surrounding repeat elements²⁴, interact with the imprinted *H19* gene locus²⁵.

We conclude that the recognition of key chromatin motifs during chromosomal interactions involves a combination of chromatin movements, chromatin fibre collisions and the stabilization of these interactions as a result of specific DNA-protein complexes and epigenetic marks. It will be important to address the functional effects of such

chromatin loops and bridges and how these features are regulated during pivotal biological processes.

Chromatin crosstalk and transcriptional activation

It has been proposed that transcriptional activation is associated with subcompartments termed transcription factories. First postulated in ref. 10, transcription factories are thought to support the simultaneous transcription of many genes, thus providing opportunities for chromatin crosstalk both in *cis* and in *trans*¹². The suggestion that transcription factories that are visualized by using antibodies directed against the active RNA polymerase II may not be functionally homogeneous²⁶ is gaining support. For example, minichromosomes containing different sets of transcriptional units uncovered at least five kinds of specialized transcription factory, according to the promoter type, the presence of introns and the type of transcribing polymerase²⁷.

How transcription factories in general, and these specialized transcription factories in particular, are formed is not known. One possibility is that transcriptional units poised for transcription attain increased mobility to explore nuclear microenvironments, eventually leading to recognition and association with a subset of previously formed transcription factories¹². If these are equipped with a key factor, an initial encounter may stabilize the interaction, eventually triggering transcription. It is also possible, however, that transcription factories are formed only after the clustering of genes. For example, the interleukin 4, 5 and 13 genes cluster in type-2 helper T cells (T_H2 cells), physically juxtaposing regulatory elements before transcriptional activation coordinated by the T_H2 -cell locus control region (LCR)²⁸. Transcriptional activation could then be initiated by the formation of transcription factories on such clustered genes.

The enhancers are likely to play a major part in these scenarios by driving the physical clustering of genes²⁸. Such complexes might involve direct communication between enhancer and promoter regions in *cis* and in *trans* to prepare for transcription by modifying chromatin marks along

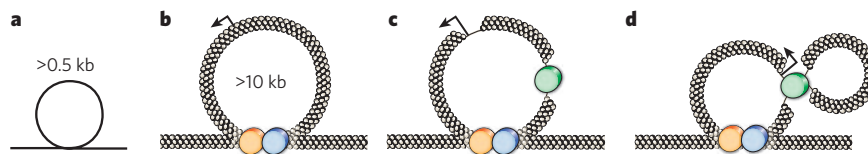


Figure 1 | Structural constraints of DNA/chromatin loop formation.

A loop containing only DNA must be larger than 0.5 kb (a), whereas a chromatin loop needs to be more than 10 kb in length (b) to form¹⁸. Upon chromatin remodelling and eviction of nucleosomes at promoter (arrow)

and/or enhancer (green sphere) regions (c), the naked DNA could provide a 'hinge' region, and thus opportunities for creating chromatin loops smaller than 10 kb (d) may arise. The orange and blue spheres represent protein complexes that organize the basic loop structure.

the chromatin fibres from the enhancer²⁹ and simultaneously increasing their mobility^{16,22}. The complexity of this process is highlighted by the demonstration that one enhancer can stochastically communicate with multiple promoters³⁰; multiple enhancers can also crosstalk with a single promoter³¹. Not mutually exclusive with this possibility is that enhancers might also operate by anchoring the transcriptional unit to a transcription factory to trigger transcriptional activation. Moreover, it is not known whether the enhancers and promoters engage in crosstalk at the time the locus is functionally incorporated into a transcription factory to modulate the efficiency of transcription. However, an analogous principle has been used to explain the efficiency of the rRNA transcription process. The promoter and terminator regions of the rRNA gene physically interact to facilitate reinitiation of transcription³². This principle ensures efficient transcription by keeping the polymerase complex in the loop, thus contributing to the fact that more than half of all RNAs in most living cells are made up of ribosomal transcripts.

Chromatin crosstalk and transcriptional silencing

The separation of euchromatic (active) and heterochromatic (inactive) domains is a common theme throughout evolution. Apart from maintaining constitutive heterochromatin at functionally essential regions, such as centromeres, this mechanism ensures stable inheritance of the lineage-specific gene expression patterns that specify various cell types. Thus, chromatin crosstalk must not traverse these boundaries unless it involves a dynamic change in transcriptional potential¹⁴. This separation can be achieved through the establishment of chromatin insulators that prevent enhancer functions from leaking inappropriately into neighbouring domains, and through the formation of chromatin barriers that prevent the silencing features of heterochromatin from inappropriately

spreading into neighbouring active domains³³. The 11-zinc-finger protein CTCF is currently the only known insulator protein in mammals³⁴ and has chromatin barrier properties³⁵. CTCF-binding sites follow the density of genes³⁶ and flank nuclear-lamina-associated heterochromatic regions⁸. As CTCF also forms complexes with proteins that may relocate chromatin fibres to nuclear subcompartments, such as the nucleolus³⁷, CTCF emerges as a key component in the functional organization of the mammalian nuclear architecture.

The 3C method³⁸ (Box 1) revealed that the CTCF-dependent insulator sites at the HS4 site at the 5' boundary of the β -globin gene³⁹, as well as at the *H19* imprinting control region (ICR)⁴⁰, form transient interactions with the chromatin fibres of the neighbouring transcriptional units. Deleting or mutating the CTCF-binding sites at the *H19* ICR led to *de novo* DNA methylation not only at the ICR but also at a key regulatory element located in *cis*⁴⁰. Similarly, targeted deletion of CTCF-binding sites within the β -globin gene locus not only disrupted long-range chromatin loops but also induced local loss of histone acetylation and gain of histone methylation³⁹. The mechanism underlying the deposition of epigenetic marks established by such chromatin crosstalk is still poorly understood. One possibility is that an interaction between CTCF and SUZ12 juxtaposes the polycomb repressive complex 2 (PRC2) with chromatin fibres interacting with the *H19* ICR to establish repressive H3K27me3 (histone H3 trimethylated at Lys 27) marks⁴¹. Interestingly, the distribution of the cohesin complex on the chromatin fibre extensively overlaps with that of CTCF-binding sites, suggesting that cohesin might contribute to the stability of such chromatin loops^{42–45}, perhaps guiding PRC2 to other CTCF-binding sites⁴⁶.

Derivatives of the 3C method that allow the identification of unknown sequences interacting with known sequences (Box 1) have revealed extensive crosstalk between the *H19* ICR²⁵ or the β -globin gene LCR⁴⁷ and the rest of the genome. At least in the case of the *H19* ICR, this 'chromosome interactome' seems to have an effect on the expression of several participating members. Maternal inheritance of mutations in the CTCF-binding sites of the *H19* ICR not only disrupted its interactions with the *Wsb1-Nf1* domain⁴⁸ and the *Osblp1a-Impact* imprinted domain²⁵ but also led to changes in the expression levels of these loci. The widely assumed function of chromatin insulators and barriers (to partition expression domains or prevent crosstalk in *cis* between euchromatin and heterochromatin) should thus be extended to include their ability to fine-tune gene expression in *trans* by means of chromatin crosstalk.

This function extends to the regulation of the X-chromosome inactivation process. Thus, CTCF-mediated interaction between the two X chromosomes in female mammals seems to be an essential part of the counting phase and inactivation process^{49,50}. X-chromosome inactivation is subsequently manifested by the creation of a repressive pocket that lacks transcription factors on the future inactive X chromosome, a process mediated by a non-coding transcript termed *Xist*⁵¹. The inactivation process is tightly linked with the recruitment of most X-linked genes into this pocket, although the cause-and-effect relationships of these events are currently not known. The repressive pocket is likely to depend on H3K27me3 marks, which are laid down by PRC2, because EZH2, a component of PRC2, interacts directly with the *Xist* RNA⁵².

There is increasing evidence that chromatin loops are also involved in polycomb-mediated gene silencing. In *Drosophila*, polycomb-mediated silencing seems to be enhanced by interactions between DNA sequences

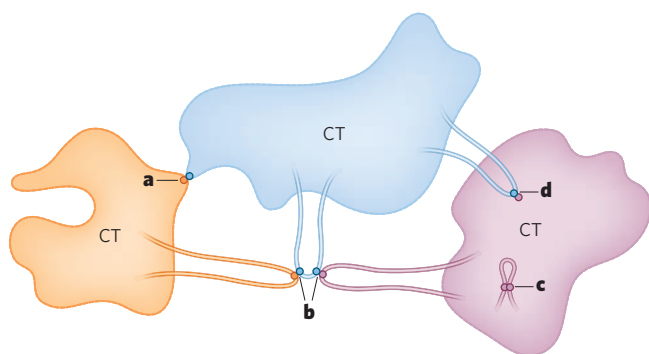


Figure 2 | Intrachromosomal and interchromosomal interactions in relation to chromosome territories.

a, An interchromosomal interaction between loci at the edge of the associated chromosome territories (CTs). b, Multiple long-range interactions with the interacting loci looping out of their CTs. This scenario is tentatively supported by the identification of up to five chromosomes simultaneously impinging on each other, as determined by using the 4C technique²⁵ (Box 1). c, An intrachromosomal interaction occurring within the CT. d, An interaction in which one locus loops into the CT of another chromosome to find its partner. The illustration does not take into consideration the dynamics of CTs. Moreover, the long-range loops and bridges may not be based on single chromatin fibres but might instead consist of thin extensions of the CT itself to reduce the potential for DNA breaks.

containing polycomb-repressor-binding elements. Indeed, all major polycomb-bound elements at the bithorax complex multigene locus display extensive chromatin loops⁵³, implying that the three-dimensional structure of chromatin plays a role in the maintenance of cellular identity. Similarly, polycomb proteins organize dynamic chromatin loops to keep the *GATA4* gene inactive in human embryonic carcinoma cells⁵⁴. However, such polycomb-dependent chromatin loops may not provide a generally applicable explanation for inactive chromatin hubs. For example, whereas the α -globin gene cluster is associated with polycomb when inactive, the β -globin gene cluster is not⁵⁵. This distinction might reflect the fact that the β -globin cluster is in a relatively gene-poor region, whereas the α -globin cluster resides in a gene-rich domain, thus demonstrating the context-dependent principles for the organization of inactive domains.

Chromatin crosstalk and nuclear architecture

The tendency towards spatial separation between active and inactive regions influences the organization of the genome within the nuclear architecture. Thus, gene-poor chromosomes are likely to be present at the nuclear periphery, whereas gene-rich chromosomes tend to occupy more internal positions. The same principle drives the organization of chromosome territories of individual chromosomes²¹. As a result, (G+C)-rich gene clusters generally displaying open chromatin structure localize preferentially in the nuclear interior, whereas (A+T)-rich constitutive heterochromatin is positioned towards the nuclear lamina and perinucleolar space⁷ (Fig. 3). The chromatin loops or chromosome territory extensions contributing to this arrangement in the nucleus might therefore reflect the formation of specialized subcompartments for gene transcription and silencing, where high-level transcription is associated with a more internal position but is not totally excluded from the nuclear periphery. In fact, many of the gene–gene interactions determined by the 3C, 4C and FISH methods can be accounted for by their co-localization to special subnuclear compartments, such as transcription factories⁹ and splicing speckles⁵⁶. However, not all nuclei display this arrangement of active and silent compartments. The structural plasticity of nuclear architecture is illustrated by the remarkable observation that it can undergo a complete reorganization in some cells. For example, rod cells in the eyes of nocturnal mammals display ‘inverted’ architecture, in which all heterochromatic portions localize to the centre of the nucleus and genes map to the nuclear periphery irrespective of their transcriptional activity⁵⁷.

This observation raises the issue of how nuclear architecture can be stably maintained and yet simultaneously allow dynamic behaviour. It is possible that these features depend on the physical properties of the nuclear environment, which — being a system containing large amounts of polymers, such as nucleic acids and protein complexes — is a classic example of macromolecular crowding⁵⁸. The physical laws operating in such systems influence the dynamics of chromatin structures and nuclear subcompartments. Non-local interactions strongly promote the compaction of chromosomes into chromosome territories without anchoring them to an immobile platform⁵⁸. The same principles favour the segregation of macromolecules into aggregates, such as transcription factories, on the basis of their shapes and affinities for each other, without restricting the exchange of their contents with the diffusible pool of the nucleoplasm⁵⁸. This raises the question of how the radial arrangement of chromosome territories is maintained in conventional nuclei and reorganized in inverted nuclei. One possibility is that gene-poor chromosomes are tethered to the nuclear lamina by an interaction between the nuclear lamina and chromatin. This scenario is supported by the observation that abrogation of lamin B1 expression in mouse embryonic fibroblasts leads to loss of chromosome-18 anchorage to the nuclear lamina⁵⁹. Similarly, a mutation in the gene encoding lamin A compromises the ability of chromosome territories to reorganize themselves when the cells leave the cell cycle⁶⁰. It will be interesting to see how inversion of nuclear architecture negotiates the interaction between lamins and heterochromatin and why the conventional architecture is more common than the inverted one. It has been suggested that the conventional architecture might have been selected for because it allows flexible chromosome arrangements and provides positional information about nuclear functions⁵⁷.

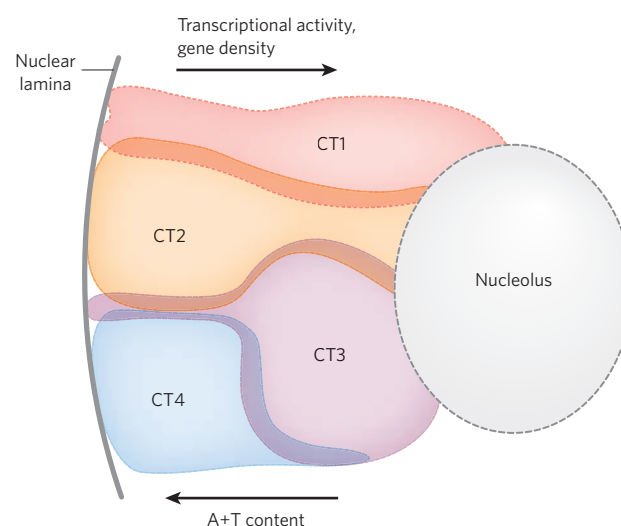


Figure 3 | Radial organization of chromosome territories within the nucleus regulates opportunities for chromatin crosstalk. The relative positions of chromosomes in an interphase nucleus depend on the proportion of genes and the A+T content. The opportunities for chromatin crosstalk between gene-rich and gene-poor regions are thus generally restricted by this organization. Hypothetical areas of chromatin communication are indicated by the patterns of overlapping CTs. Reorganization of CTs could provide new patterns of chromosomal interactions. The presence of the nucleolus and many other subnuclear compartments (not shown) may provide additional opportunities for the formation of chromatin loops and bridges.

A few examples hint at the possibility that the interplay between nuclear subcompartmentalization and the formation of chromatin loops and bridges could indeed increase the sophistication of transcriptional regulation by diversifying transcriptional states and influencing the kinetics of gene transcription^{22,61}. The ligand-induced physical clustering of a specific subset of transcriptional units bound to oestrogen receptor- α (ER- α ; also known as ESR1) illustrates how the formation of interchromosomal interactions before transcription is likely to facilitate coordinated and efficient transcription²². Although both the ER- α -bound interacting units and the non-interacting ones display chromatin structure that is permissive for transcription, only the interacting loci become relocated — by the action of the histone lysine demethylase LSD1 (also known as KDM1) — to interchromosomal granules that contain transcription elongation and splicing factors²².

Nuclear architecture could also be reflected in the replication process. Whether a region replicates during the early, middle or late S phase of the cell cycle strongly correlates with its position within the nuclear architecture, perhaps as a result of the various chromatin conformations and their availability to replication factors⁶². Replication is proposed to take place in replication factories that might harbour up to a dozen simultaneously replicating sequences. As large regions (domains of up to several million base pairs) need to be replicated within just a few hours, the coordination of origin firing over large distances is likely to involve chromatin crosstalk⁶³. It has been proposed that, to achieve this, ‘licensed’ origins coalesce in *cis* before initiation of DNA replication. As a result of such interactions, licensed origins might be able to coordinate the firing and, hence, the timing of replication of large subchromosomal domains¹⁵. Because the timing of replication can influence the potential for transcription, interactions between licensed replication origins might govern the pattern of gene expression in the subsequent cell cycle, and thus link positional information within the nuclear architecture to either propagation or reprogramming of epigenetic states during cell division.

Noise and order in chromatin crosstalk

Although it is generally accepted that chromatin fibres interact with each other in mammalian cells, what this means in functional terms is much less clear. The large number of physical interactions captured by 3C, 4C and FISH methods, as well as evident from the diversity

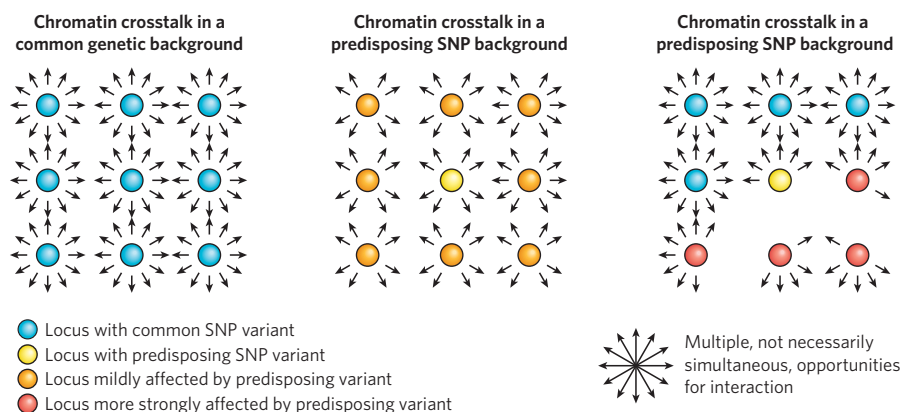


Figure 4 | Genetic background may influence the expressivity of the genome through chromosome loops and/or bridges. Chromosomal networks, defined by two or more nodes of interaction, may coordinate and fine-tune transcription (left panel). Allelic variants may stabilize or antagonize such networks, modulating gene expression patterns. The middle and

right panels illustrate different scenarios of perturbed chromatin crosstalk (compared with the advantageous variant schematically depicted in the left panel) as potentially resulting from disease-predisposing SNPs. The severity and character of disease phenotype may depend on the number of genes affected, the extent of change in their expression and their function.

of chimaeric transcripts identified⁶⁴ (see page 206), implies that the nuclear architecture has a dynamic nature with a high level of stochastic collisions between chromatin fibres that do not necessarily influence genomic functions. Indeed, it is not always clear whether the juxtaposition of distal regulatory sequences and genes represents processes that are directly involved in gene regulation and are thus causal or, instead, merely represents consequences of such regulation. An example in which the cause-and-effect relationship has been elucidated is provided by the observation that a physical interaction between the interferon- γ gene promoter on one chromosome and the T_H2 -cell LCR that coordinates the expression of interleukin genes on another chromosome fine-tunes the kinetics of transcriptional activation of the interferon- γ gene upon T-cell differentiation⁶¹.

Another example is the restriction of interferon- β gene expression to a particular environmental context, for example viral infection. The stochastic allelic expression of the interferon- β gene requires nuclear factor- κ B, which is a rate-limiting factor for the assembly of the interferon- β enhanceosome. Viral exposure triggers the juxtaposition of Alu repeat segments from different chromosomes with the interferon- β locus. As these Alu elements carry nuclear factor- κ B, their interaction allows the formation of the enhanceosome and thus transcriptional activation of the interferon- β gene²³. Despite the difficulties in proving cause and effect, these examples convincingly illustrate how chromatin crosstalk can functionally increase the adaptive plasticity of the cell exposed to the changing microenvironment.

Although noise in chromatin crosstalk would be expected to be largely non-functional, if stabilized it might contribute to phenotypic diversity, for example by establishing and/or maintaining stochastic patterns of monoallelic expression⁶⁵. It is currently unknown, however, the way in which stable interactions between distant elements are orchestrated, especially in the context of the chromosome territories. At least part of the solution to this mystery may lie in the observation that the relative positions of the chromosome territories are subject to developmental regulation⁶⁶, influencing the probability of interactions in *trans* in cell-type-specific ways. Whether cohesin is the stabilizing factor, as was recently shown for chromatin fibre interactions in *cis*⁴⁵, remains to be determined. Irrespective of what factor or factors are involved, a key issue is how these interactions are specified, for example among thousands of binding sites in the instance of cohesin.

Perspectives

Without doubt, an emerging major challenge in chromatin biology is to unravel the mechanism(s) of interactions between chromosomes in three dimensions and to map and understand the influence of this interactome on the expressivity of the genome. Real progress in the field will depend on the development of new strategies and technologies that

more precisely allow the definition of the cause-and-effect relationship between chromosomal interactions and genomic functions. Although real-time imaging of interactions between different loci is technically feasible and is essential for understanding how chromatin mobility is regulated in relation to a biological process, it has low resolution and cannot readily be used to screen genome-wide patterns of chromatin interactions. Other limitations apply to the 3C, 4C and 5C techniques, which require large populations of cells for analysis and hence do not readily advance our understanding of the dynamics of chromatin crosstalk. A new strategy is needed, therefore, to address the forces driving higher-order chromatin folding and to observe simultaneous co-localization events in relation to nuclear subcompartments at high resolution in individual cells. This would allow comparisons testing of a range of variables, such as how an interaction responds to the micro-environment, the three-dimensional position of the interaction, and when in the cell cycle the interaction occurs. Ideally, this should allow the identification of the molecular factors and chromatin marks participating in the interaction, enabling us to understand the phenotypic read-out effect of the interaction.

Addressing these issues may ultimately yield new perspectives on how chromatin crosstalk influences human diseases. For example, we may uncover why genome-wide association studies of complex diseases often map to gene deserts⁶⁷. As chromatin loop formation has been documented to be sensitive to particular combinations of sequence polymorphisms⁶⁸, one possibility is that particular sets of single-nucleotide polymorphisms (SNPs) may influence communication between different parts of the genome by inducing or abolishing loop formation. Figure 4 shows two potential scenarios for how a particular combination of SNPs within a gene desert might generate pleiotropic changes in gene function elsewhere in the genome, either through the formation of disadvantageous chromosomal interactions or through the loss of advantageous patterns of interactions. Furthermore, chromatin crosstalk can be linked to misregulation of nuclear processes, as it may provide a platform for chromosomal translocation events between genes that are frequently transcribed in the same transcription factory^{69,70}. For these and many other reasons, it may be useful to integrate the concept of chromosome interactomes when exploring the genetic and/or epigenetic background of complex diseases, including cancer.

1. Wilson, E. *The Cell in Development and Heredity* (Macmillan, 1928).
2. Duncan, I. W. Transvection effects in *Drosophila*. *Annu. Rev. Genet.* **36**, 521–556 (2002).
3. Cook, P. R. The organization of replication and transcription. *Science* **284**, 1790–1795 (1999).
4. Misteli, T. Nuclear order out of chaos. *Nature* **456**, 333–334 (2008).
5. Kaiser, T. E., Intine, R. V. & Dundr, M. De novo formation of a subnuclear body. *Science* **322**, 1713–1717 (2008).
6. McStay, B. & Grummt, I. The epigenetics of rRNA genes: from molecular to chromosome biology. *Annu. Rev. Cell Dev. Biol.* **24**, 131–157 (2008).

7. Lanctot, C., Cheutin, T., Cremer, M., Cavalli, G. & Cremer, T. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nature Rev. Genet.* **8**, 104–115 (2007).
8. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
9. Osborne, C. S. *et al.* Active genes dynamically colocalize to shared sites of ongoing transcription. *Nature Genet.* **36**, 1065–1071 (2004).
10. Jackson, D. A., Hassan, A. B., Errington, R. J. & Cook, P. R. Visualization of focal sites of transcription within human nuclei. *EMBO J.* **12**, 1059–1065 (1993).
11. Faro-Trindade, I. & Cook, P. R. Transcription factories: structures conserved during differentiation and evolution. *Biochem. Soc. Trans.* **34**, 1133–1137 (2006).
12. Fraser, P. & Bickmore, W. Nuclear organization of the genome and the potential for gene regulation. *Nature* **447**, 413–417 (2007).
13. Hassan, A. B. & Cook, P. R. Visualization of replication sites in unfixed human cells. *J. Cell Sci.* **105**, 541–550 (1993).
14. Chakalova, L., Debrand, E., Mitchell, J. A., Osborne, C. S. & Fraser, P. Replication and transcription: shaping the landscape of the genome. *Nature Rev. Genet.* **6**, 669–677 (2005).
15. Göndör, A. & Ohlsson, R. Replication timing and epigenetic reprogramming of gene expression: a two-way relationship? *Nature Rev. Genet.* **10**, 269–276 (2009).
16. Soutoglou, E. & Misteli, T. Mobility and immobility of chromatin in transcription and genome stability. *Curr. Opin. Genet. Dev.* **17**, 435–442 (2007).
17. Dostie, J. *et al.* Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16**, 1299–1309 (2006).
18. Rippe, K. Making contacts on a nucleic acid polymer. *Trends Biochem. Sci.* **26**, 733–740 (2001).
19. Li, Q., Barkess, G. & Qian, H. Chromatin looping and the probability of transcription. *Trends Genet.* **22**, 197–202 (2006).
20. Ishihara, K., Oshimura, M. & Nakao, M. CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol. Cell* **23**, 733–742 (2006).
21. Cremer, T. & Cremer, C. Rise, fall and resurrection of chromosome territories: a historical perspective. Part II. Fall and resurrection of chromosome territories during the 1950s to 1980s. Part III. Chromosome territories and the functional nuclear architecture: experiments and models from the 1990s to the present. *Eur. J. Histochem.* **50**, 223–272 (2006).
22. Hu, Q. *et al.* Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules. *Proc. Natl Acad. Sci. USA* **105**, 19199–19204 (2008).
This report demonstrates that chromatin crosstalk involves directed rapid movement of interacting loci and that this process is frequently accompanied by reorganization of the chromosome territories.
23. Apostolou, E. & Thanos, D. Virus infection induces NF- κ B-dependent interchromosomal associations mediating monoallelic IFN- β gene expression. *Cell* **134**, 85–96 (2008).
This report shows that the organization of the IFN β enhanceosome requires the juxtaposition of one or more sequences from other chromosomes that contain NF- κ B-bound Alu repeats.
24. Walter, J., Hutter, B., Khare, T. & Paulsen, M. Repetitive elements in imprinted genes. *Cytogenet. Genome Res.* **113**, 109–115 (2006).
25. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genet.* **38**, 1341–1347 (2006).
26. Bartlett, J. *et al.* Specialized transcription factories. *Biochem. Soc. Symp.* **73**, 67–75 (2006).
27. Xu, M. & Cook, P. R. Similar active genes cluster in specialized transcription factories. *J. Cell Biol.* **181**, 615–623 (2008).
28. Cai, S., Lee, C. C. & Kohwi-Shigematsu, T. SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nature Genet.* **38**, 1278–1288 (2006).
29. Vernimmen, D., De Gobbi, M., Sloane-Stanley, J. A., Wood, W. G. & Higgs, D. R. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J.* **26**, 2041–2051 (2007).
30. Tsytsykova, A. V. *et al.* Activation-dependent intrachromosomal interactions formed by the TNF gene promoter and two distal enhancers. *Proc. Natl Acad. Sci. USA* **104**, 16850–16855 (2007).
31. Deschenes, J., Bourdeau, V., White, J. H. & Mader, S. Regulation of *GREB1* transcription by estrogen receptor α through a multipartite enhancer spread over 20 kb of upstream flanking sequences. *J. Biol. Chem.* **282**, 17335–17339 (2007).
32. Nemeth, A., Guibert, S., Tiwari, V. K., Ohlsson, R. & Langst, G. Epigenetic regulation of TTF-I-mediated promoter-terminator interactions of rRNA genes. *EMBO J.* **27**, 1255–1265 (2008).
33. Bushey, A. M., Dorman, E. R. & Corces, V. G. Chromatin insulators: regulatory mechanisms and epigenetic inheritance. *Mol. Cell* **32**, 1–9 (2008).
34. Ohlsson, R., Renkawitz, R. & Lobanov, V. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet.* **17**, 520–527 (2001).
35. Kim, J. H. *et al.* Human gamma-satellite DNA maintains open chromatin structure and protects a transgene from epigenetic silencing. *Genome Res.* **19**, 533–544 (2009).
36. Kim, T. H. *et al.* Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**, 1231–1245 (2007).
37. Yusufzai, T. M., Tagami, H., Nakatani, Y. & Felsenfeld, G. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol. Cell* **13**, 291–298 (2004).
38. Dekker, J. The three 'C's of chromosome conformation capture: controls, controls, controls. *Nature Methods* **3**, 17–21 (2006).
39. Splinter, E. *et al.* CTCF mediates long-range chromatin looping and local histone modification in the β -globin locus. *Genes Dev.* **20**, 2349–2354 (2006).
40. Kurukuti, S. *et al.* CTCF binding at the *H19* imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to *Igf2*. *Proc. Natl Acad. Sci. USA* **103**, 10684–10689 (2006).
41. Li, T. *et al.* CTCF regulates allelic expression of *Igf2* by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. *Mol. Cell Biol.* **28**, 6473–6482 (2008).
42. Pareilho, V. *et al.* Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132**, 422–433 (2008).
43. Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).
44. Stedman, W. *et al.* Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and *H19/Igf2* insulators. *EMBO J.* **27**, 654–666 (2008).
45. Hadjur, S. *et al.* Cohesins form chromosomal cis-interactions at the developmentally regulated *IFNG* locus. *Nature* **460**, 410–413 (2009).
46. Göndör, A. & Ohlsson, R. Chromatin insulators and cohesins. *EMBO Rep.* **9**, 327–329 (2008).
47. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genet.* **38**, 1348–1354 (2006).
48. Ling, J. Q. *et al.* CTCF mediates interchromosomal colocalization between *Igf2/H19* and *Wsb1/Nf1*. *Science* **312**, 269–272 (2006).
49. Masui, O. & Heard, E. RNA and protein actors in X-chromosome inactivation. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 419–428 (2006).
50. Xu, N., Donohoe, M. E., Silva, S. S. & Lee, J. T. Evidence that homologous X-chromosome pairing requires transcription and Ctfp protein. *Nature Genet.* **39**, 1390–1396 (2007).
51. Chaumeil, J., Le Baccon, P., Wutz, A. & Heard, E. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev.* **20**, 2223–2237 (2006).
52. Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756 (2008).
53. Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F. & Orlando, V. Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nature Cell Biol.* **9**, 1167–1174 (2007).
54. Tiwari, V. K. *et al.* PcG proteins, DNA methylation, and gene repression by chromatin looping. *PLoS Biol.* **6**, e306 (2008).
55. Garrick, D. *et al.* The role of the polycomb complex in silencing α -globin gene expression in nonerythroid cells. *Blood* **112**, 3889–3899 (2008).
56. Brown, J. M. *et al.* Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J. Cell Biol.* **182**, 1083–1097 (2008).
57. Solovei, I. *et al.* Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell* **137**, 356–368 (2009).
This report demonstrates that, contrary to expectations, the nuclear architecture in the rod cells of the eye can undergo dramatic reorganization to adapt to a new function associated with nocturnal mammals.
58. Richter, K., Nessling, M. & Lichter, P. Macromolecular crowding and its potential impact on nuclear function. *Biochim. Biophys. Acta* **1783**, 2100–2107 (2008).
59. Shimi, T. *et al.* The A- and B-type nuclear lamin networks: microdomains involved in chromatin organization and transcription. *Genes Dev.* **22**, 3409–3421 (2008).
60. Mehta, I. S., Elcock, L. S., Amira, M., Kill, I. R. & Bridger, J. M. Nuclear motors and nuclear structures containing A-type lamins and emerin: is there a functional link? *Biochem. Soc. Trans.* **36**, 1384–1388 (2008).
61. Spiliakakis, C. G., Lalioti, M. D., Town, T., Lee, G. R. & Flavell, R. A. Interchromosomal associations between alternatively expressed loci. *Nature* **435**, 637–645 (2005).
62. Aladjem, M. I. Replication in context: dynamic regulation of DNA replication patterns in metazoans. *Nature Rev. Genet.* **8**, 588–600 (2007).
63. Hiratani, I. *et al.* Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* **6**, e245 (2008).
64. Unneberg, P. & Claverie, J. M. Tentative mapping of transcription-induced interchromosomal interaction using chimeric EST & mRNA data. *PLoS ONE* **2**, e254 (2007).
65. Ohlsson, R. Widespread monoallelic expression. *Science* **318**, 1077–1078 (2007).
66. Parada, L. A., McQueen, P. G. & Misteli, T. Tissue-specific spatial organization of genomes. *Genome Biol.* **5**, R44 (2004).
67. Ghousaini, M. *et al.* Multiple loci with different cancer specificities within the 8q24 gene desert. *J. Natl Cancer Inst.* **100**, 962–966 (2008).
68. Steidl, U. *et al.* A distal single nucleotide polymorphism alters long-range regulation of the *PU.1* gene in acute myeloid leukemia. *J. Clin. Invest.* **117**, 2611–2620 (2007).
69. Branco, M. R. & Pombo, A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol.* **4**, e138 (2006).
70. Osborne, C. S. *et al.* Myc dynamically and preferentially relocates to a transcription factory occupied by *Igh*. *PLoS Biol.* **5**, e192 (2007).

Acknowledgements We most gratefully acknowledge I. Solovei, B. Joffe, P. Cook, G. Klein and E. Heard for discussions. This work was supported by the Swedish Science Research Council, the Swedish Cancer Research Foundation, the Swedish Pediatric Cancer Foundation, the Lundberg Foundation, and HEROIC and CHILL (European Union integrated projects).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence should be addressed to A.G. (anita.gondor@ki.se) or R.O. (rolf.ohlsson@ki.se).

Molecular networks as sensors and drivers of common human diseases

Eric E. Schadt¹

The molecular biology revolution led to an intense focus on the study of interactions between DNA, RNA and protein biosynthesis in order to develop a more comprehensive understanding of the cell. One consequence of this focus was a reduced attention to whole-system physiology, making it difficult to link molecular biology to clinical medicine. Equipped with the tools emerging from the genomics revolution, we are now in a position to link molecular states to physiological ones through the reverse engineering of molecular networks that sense DNA and environmental perturbations and, as a result, drive variations in physiological states associated with disease.

Our understanding of common human diseases and how best to treat them is hampered by the complexity of the human system in which they are manifested. Unlike simple Mendelian disorders, in which highly expressive, highly penetrant mutations make it possible to identify the causal genes within families in which traits associated with the disorders segregate¹, common human diseases originate from a more complex interplay between constellations of changes in DNA (both rare and common variations) and a broad range of factors such as diet, age, gender and exposure to environmental toxins.

These complex arrays of interacting factors are thought to affect entire network states that in turn increase or decrease the risk of disease or affect disease severity. In the context of common human diseases, the disease states can be considered emergent properties of molecular networks², as opposed to the core biological processes associated with a disease being driven by responses to changes in a small number of genes. Integrating large-scale, high-dimensional molecular and physiological data holds promise not only for defining the molecular networks that directly respond to genetic and environmental perturbations that associate with disease but also for causally associating such networks with the physiological states associated with disease. Given what must be considered a deluge of data of many different types flooding life sciences and biomedical research today, including genome-wide single-nucleotide polymorphism (SNP) genotyping data, whole-genome transcription data, next-generation DNA sequencing data, RNA sequencing data, chromatin immunoprecipitation (ChIP) sequencing data and image data, it is now time to begin addressing how these large-scale, high-dimensional data sets can be integrated to better understand the molecular networks underlying physiological states associated with disease. Here, I review the progress made over the past few years to integrate DNA variation, molecular profiling and clinical data collected in populations in order to construct causal probabilistic networks of disease, providing a more comprehensive view of disease than can be achieved by examining the different data dimensions on their own. Particular attention is paid to describing how the predictive networks produced from this type of integrative modelling can help link molecular states to physiological ones, providing an alternative path for understanding how molecular states drive complex disease processes.

GWAS provide insights into human diseases

Roughly three billion nucleotides make up the human genome, so the number of nucleotide changes that can affect the activities of genes is

effectively infinite with respect to our ability to determine the effects of combinations of such changes experimentally. Therefore, exploiting naturally occurring DNA variation in human populations is among the most attractive approaches to inferring the constellation of genes that affect disease risk. For most diseases, changes in DNA that correlate with disease can be inferred as tagging or directly representing causal components of disease. Therefore, DNA variation directly elucidates disease aetiology and is extremely useful (Fig. 1a). Genome-wide association studies (GWAS) are now well proven to uncover genetic loci that affect disease risk or progression³.

The emergence of technologies capable of characterizing DNA variation systematically over the entire genome and in whole populations has revolutionized our ability to apply GWAS approaches to many human diseases, with more than 200 loci now identified and highly replicated for Crohn's disease⁴, type 2 diabetes⁵, serum lipid levels^{6,7}, prostate cancer^{8,9}, age-related macular degeneration^{10,11}, obesity¹² and more than 50 other human diseases³. By comparing the frequencies of genetic variants between individuals with and without disease, or by directly testing for correlations between a quantitative disease trait and genotypes at a given locus, GWAS can lead directly to the causal variants of disease or to variants that are in strong linkage disequilibrium with variants of disease. Therefore, the power of approaches such as GWAS lies in their ability to identify the genetic causes of disease, which can be used to predict disease risk and to elucidate signalling pathways associated with disease, information that is of use in drug discovery.

Integrative genomics and disease networks

GWAS have uncovered many genetic loci that associate with human diseases, but two fundamental limitations have hampered our ability to translate these results into clinically useful predictors of disease and drug targets. First, the genetic loci associated with disease generally explain very little of the disease risk. The odds of having a risk genotype at a particular disease locus given that you have the disease, divided by the odds of having a risk genotype given that you do not have the disease, are typically less than 1.5 (ref. 3). Second, the SNP-trait associations alone do not necessarily lead directly to the identification of the causal gene(s), much less elucidate the context in which the causal gene(s) operates^{3,13,14}. Understanding the biological context in which a given causal gene for disease operates is a necessary step in identifying the best drug targets^{15,16}.

¹Pacific Biosciences, 1505 Adams Drive, Menlo Park, California 94025, USA.

Interestingly, in the span of just a few years, the realization that tractable drug targets and clinically useful biomarkers of disease are not immediately apparent from GWAS data has, for some, reduced enthusiasm for the GWAS approach^{17–19}. However, given that variations in DNA do not on their own directly impact on physiological states associated with disease, there is the potential to enhance our understanding of GWAS data by layering in a hierarchy of phenotypes that define the molecular and physiological states associated with disease^{13,14,20–22}. Because variations in DNA more proximally (relative to disease states) induce changes in molecular states that in turn drive variations in physiological states associated with disease, incorporating such data can allow the identification of causal genes and the broader biological context in which they operate. Therefore, elucidating changes in molecular states that more directly respond to changes in DNA and that in turn influence disease has the potential to fill in the gaps left by GWAS.

In fact, the advances made in mapping DNA loci for diseases have occurred simultaneously with the mapping of DNA loci for molecular traits such as transcript abundances^{13,14,20,22–24}. Identifying the RNAs that mediate the flow of information from DNA to disease is of particular interest in this context, given that, because it is transcribed directly from a DNA template, RNA is the most proximal non-DNA species of all molecular entities in the cell. In studies that seek to map genetic loci that affect RNA levels, SNP genotypes are tested for association with tens of thousands of RNA traits scored simultaneously in population samples. A number of such studies have demonstrated that the amount of variation in RNA levels explained by a given genetic locus can often be greater than 50% (refs 13, 14, 22 and 24). In addition, family-based studies of the genetics of RNA levels in multiple tissues have estimated that a majority of RNA traits on average have a genetic variance component of 30% (ref. 13). The mapping of genetic loci for molecular traits is not constrained only to RNA levels. Any molecular species that can be reasonably well measured (for example protein or metabolite levels) is amenable to genetic mapping and can complement genetic mapping for RNA traits²⁵. Mapping studies involving RNA traits are not without significant analysis issues. The large number of RNA traits and markers that can be tested demands that significance levels for association be rigorously adjusted to control for false-discovery rates¹⁴.

Molecular traits controlled by genetic loci associated with disease can be treated as intermediate phenotypes of disease and thus elucidate the molecular networks underlying disease. This can aid in the interpretation of GWAS data by identifying genes whose RNA levels associate with genetic loci that also associate with disease^{6,13,14,20,26,27}. Furthermore, these data can be treated more formally to infer causal relationships between molecular traits and disease states^{2,21,28,29}, a process that has been shown to aid in the identification of genes or specific isoforms of genes corresponding to loci identified in the GWAS^{14,20,30} (Fig. 1b). One of the central issues related to the use of RNA traits to enhance identification of genes in genomic regions associated with disease is assessing whether a given locus is jointly associated with disease and RNA levels, or whether two closely linked loci control the RNA levels and disease independently^{14,21}. Formal statistical procedures that examine the joint probabilities for the genotype, RNA and disease data can be applied to establish whether RNA levels and disease are related in either an independent relationship or a causal or reactive relationship^{2,21,28,29}.

The introduction of molecular traits can enhance the interpretation of GWAS results by placing them in a broader biological context that may support the identification of disease-susceptibility genes and more generally elucidate networks (Box 1) that define the biological processes associated with disease¹⁴. One of the more intriguing examples of this approach was the identification of three candidate susceptibility genes (*SORT1*, *CELSR2* and *PSRC1*) for cardiovascular disease and lipid levels^{2,7}, where the disease-associated and lipid-associated SNPs were also significantly associated with the liver expression of the three candidate genes, which were physically located near the disease-associated SNP. These genes were also supported as causal for low-density-lipoprotein cholesterol levels in a previously described experimental mouse cross². Furthermore, all three genes were found to be connected in liver gene

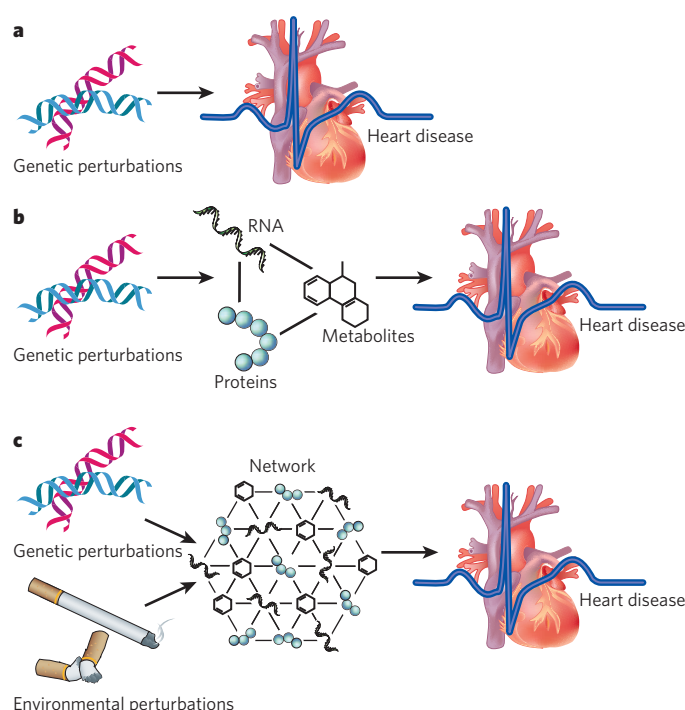


Figure 1 | Hierarchy of causal relationships. **a**, Classic genetic association approaches seek to identify variations in DNA that correlate with disease state or with quantitative traits associated with disease. The attraction of this approach is the identification of the genetic causes of disease. **b**, Changes in DNA on their own do not lead to disease but, instead, lead to changes in molecular traits that go on to affect disease risk. By layering in molecular phenotypes as intermediate phenotypes, causal relationships between genes and disease can be established directly. **c**, Disease gene networks sense constellations of genetic and environmental perturbations. Therefore, a more realistic model is one in which constellations of genetic and environmental perturbations affect molecular states of networks that in turn affect disease risk.

networks that were constructed from mouse and human liver samples and in which the constituent genes were enriched for in a previously described macrophage-enriched metabolic network associated with a number of processes related to immune function and inflammation^{2,13,14}.

Disease networks respond to disease loci

Identifying genetic loci that associate with disease and intermediate molecular phenotypes that respond more proximally to these loci and in turn cause disease are excellent first steps to uncovering the drivers of disease. However, the view of disease becoming clear from the large-scale genomic studies is that common forms of disease are emergent properties of networks whose states are affected by a complex interaction of genetic and environmental factors. To understand the behaviour of any one gene in the context of human disease, individual genes must be understood in the context of molecular networks that define the disease states. In fact, several studies have now shown that for single diseases or traits such as height, tens or even hundreds of genes may be involved but may not be randomly distributed with respect to biological function.

For example, sequencing of DNA from tumour samples found scores of genes affected by rare variations that influence cancer risk and progression. The genes affected were shown to be significantly more likely to belong to pathways known to be involved in tumorigenesis or tumour progression than was the case for the set of all genes that were resequenced as part of this study^{31,32}. In a separate study, my research group identified a macrophage-enriched metabolic network (MEMN) that in mice was strongly indicated to be causal for a number of metabolic-disease traits². The same network was not only found to be associated with metabolic traits and conserved in human populations but also to be enriched for DNA variations near

these genes that are associated with obesity, suggesting that hundreds or thousands of genes may subtly affect obesity risk³⁸. Constructing networks that underlie core biological processes associated with disease makes it possible to identify the functional units that respond to genetic perturbations and then in turn affect disease risk (Fig. 1c). In this way, any given gene can be studied in the context of many different networks to learn whether one or more of the networks in which a given gene operates influences physiological states associated with the disease. Such mappings not only allow the identification of causal relationships among genes and between genes and more complex traits such as disease^{2,21,29} but also more generally allow the construction of predictive gene networks^{2,33}.

Before this can be achieved, however, we must integrate the diverse data necessary to construct the gene networks. There have been a number of recent advances in the construction of networks capable of predicting complex system behaviour. Examining the action of many genes simultaneously in populations segregating common disease traits has led to the identification of whole gene networks that both define disease at the molecular level and drive the onset and progression of disease^{2,13,14,31–35}. The construction of these networks allows the identification of the functional units of the system underlying physiological states^{2,29,34,36,37}.

Networks generally provide a convenient framework for exploring the context within which single genes operate (Box 1). Networks are simply graphical models that comprise nodes and edges and are convenient for visualizing complex mathematical models that describe how variables of a system associate with one another in different contexts of interest. For gene networks associated with biological systems, the nodes in the network typically represent genes, gene products or other important molecular entities, and an edge between any two nodes indicates a relationship between the corresponding genes, gene products or other molecular entities. For example, an edge between two genes may indicate that the corresponding expression traits are correlated^{33,38}, that the corresponding proteins interact³⁹ or that changes in the activity of one gene lead to changes in the activity of the other²¹. Interaction, or association, networks, which have recently become widely used in the biological community, are formed by considering only pairwise relationships between genes, including protein interactions⁴⁰ and co-expression relationships^{37,41}.

Interaction networks allow the identification of subnetworks (coherent gene modules) corresponding to the functional units of a living system^{2,29,36,37,42,43}. Increasing evidence suggests that these functional units are directly linked to physiological states, defining in humans the molecular states that lead to physiological states associated with disease. Genetic perturbations that associate with disease have been shown to act through these functional units by altering the corresponding network state. The networks therefore can serve as an organizing framework for causal perturbations that lead to disease. That is, networks sense variations in the

genome, in the methylome and in the environment more generally, given that these different types of variation affect the function of the proteins or the expression levels of the genes or proteins constituting these networks, thus altering their states. In this way, the network more maximally captures, or senses, these different sources of variation and, as a result, induces changes in physiological states associated with disease (Fig. 1c).

Although there is now an extensive literature on the construction and application of interaction networks to elucidate the complexity of disease, these methods are typically applied to gene expression data alone and therefore do not strictly reflect causal relationships among gene expression traits or between expression traits and disease. Probabilistic causal networks represent an alternative approach capable of integrating multiple types of data and inferring from these data whether two or more genes are causally connected to each other or to disease traits. Bayesian network-reconstruction methods are one of the more common approaches of this sort. They provide an elegant way of incorporating diverse data pertaining to causal relationships, such as DNA variation, gene expression, protein interaction, DNA–protein binding, and proteomic and, more recently, metabolomic data. Recent work has demonstrated that by considering these types of data simultaneously, it is possible to construct networks that are able to predict future states of the representative system^{33,44}. The construction of networks in which the relationships between genes can be understood from the standpoint of causal control is one of the ultimate aims in life sciences and biomedical research, as an understanding of predictive gene networks can lead directly to drug targets and biomarkers of disease^{15,16,45}.

The MEMN is an example of a causal network constructed by integrating different data types. The MEMN was identified from liver and adipose gene expression data generated in mouse and human populations segregating metabolic-disease phenotypes. From the resultant tissue gene networks, the MEMN was identified as strongly conserved between tissues, between sexes and between species, and was strongly associated with metabolic traits related to obesity, diabetes and heart disease^{2,13}. It was also observed to respond to variations in DNA that are associated with disease traits¹³. A statistical procedure²¹ was applied to infer whether the MEMN was responding to the DNA changes and causing variations in the metabolic traits as a result or whether it was responding to changes in the metabolic traits induced by the DNA changes. The MEMN was strongly indicated to be causal for all of the obesity, diabetes and heart-disease traits scored in an experimental mouse population.

Biological processes represented in the MEMN supported the idea of macrophages as a key driver of disease pathogenesis, consistent with recent evidence that chronic inflammation is a key feature of obesity^{2,46}. Importantly, the mouse MEMN was highly conserved in humans, in

Box 1 | Gene networks

Cells comprise many tens of thousands of proteins, metabolites, RNAs and DNAs, all interacting in complex ways. In turn, complex biological systems comprise many types of cell operating within and between the many types of tissue that make up different organ systems, all of which interact in complex ways to give rise to a vast array of phenotypes that manifest themselves in living systems. Modelling the extent of such relationships between molecular entities, between cells, and between organ systems is a daunting task. Networks are a convenient framework in which to represent the relationships among these different variables. In the context of biological systems, a network can be viewed as a graphical model that represents relationships among DNAs, RNAs, proteins, metabolites and higher-order phenotypes such as disease state. In this way, networks provide a way to visualize extremely

large-scale, complex relationships among molecular and higher-order phenotypes in any given context. In this Review, I am interested in networks that represent relationships among molecular entities in a living system, as determined empirically in populations of individuals.

In this context, biological networks comprise nodes, which represent molecular entities that are observed to vary in the population under study (for example DNA variations, RNA levels, protein states or metabolite levels). Edges between the nodes represent relationships between the molecular entities, and these edges can either be directed, indicating a cause–effect relationship, or undirected, indicating an association or interaction. For example, a DNA node in the network representing a given locus that varies in a population of interest may be connected

to a transcript-abundance trait, indicating that changes at the particular DNA locus induce changes in the levels of the transcript. The potentially millions of such relationships represented in a network define the overall connectivity structure, or topology, of the network. Any realistic network topology will necessarily be complicated and nonlinear from the standpoint of the more classic biochemical pathway diagrams presented in text books and pathway databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database⁵⁴. The more classic pathway view represents molecular processes on an individual level, whereas networks represent global (population-level) metrics describing variations between individuals in a population of interest; these variations in turn define the coherent biological processes in the tissue or cells associated with the network.

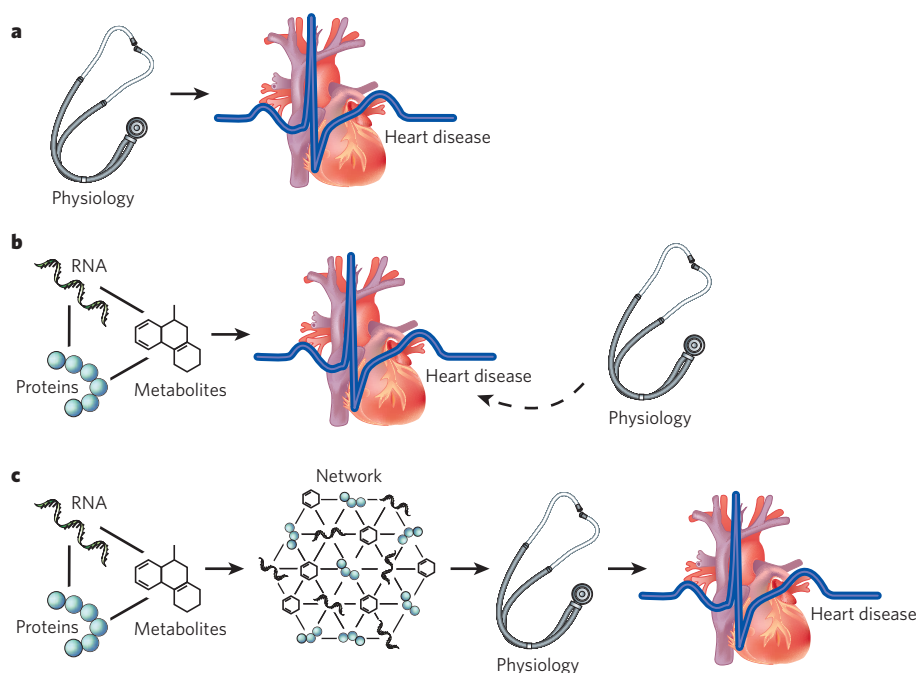


Figure 2 | Linking molecular biology to physiology through molecular networks. **a**, Before the molecular biology revolution, disease was studied primarily in the context of physiology. **b**, As a result of the molecular biology revolution, physiology has played a less prominent role in the study of the molecular bases of disease, given the reductionist push to associate molecular changes in a given gene (affecting protein levels, activity or function) directly with changes in disease states. **c**, The complexity of molecular biology — given the ability to monitor DNA variation, RNA variation, metabolite variation and protein variation in populations on a comprehensive scale — has driven a systems view of disease, in which networks of interacting molecular entities are constructed to define physiological states of the system associated with disease. In this way, the molecular networks allow a direct link between molecular biology and clinical medicine by connecting molecular biology to physiology.

whom it was also indicated to be causal for metabolic traits. A number of genes in the MEMN were predicted to be causal for metabolic-disease traits. This has now been experimentally verified, and the genes have been shown to be involved in complex feedback control, with many of them indicated and confirmed to be causal for each other^{2,14,28,29}.

Linking molecular and physiological states

The identification of the MEMN as a key driver of metabolic disease highlights several important features of the network approach to understanding disease that have implications for drug discovery: first, the network analyses revealed hundreds of disease-causing genes acting together in coherent networks; second, within a given network supported as being causal for disease, perturbing individual genes supported as being causal for disease affected the state of the network; and, third, DNA and other sources of variation in one species can be used to construct disease networks that are relevant in a second species and that act as sensors for many sources of variation (for example genetic, epigenetic and environmental sources) and in turn modulate physiological traits associated with disease. These features taken together suggest that networks such as the MEMN underlie or define the physiological states associated with disease. The data further suggest that highly efficacious treatments of diseases such as obesity might not be achieved by targeting single genes, at least not without taking into account the role of an individual gene in the network^{15,16}.

Core subnetworks associated with disease provide a path directly linking molecular biology to physiology, and it is this link that may ultimately lead to a more significant clinical impact (Fig. 2). Networks have now been modelled both within and between multiple tissues that are relevant to disease. The identification of subnetworks interacting between islet, adipose, liver, muscle and brain tissues has highlighted the importance of using a network framework directly to model physiological states associated with diabetes³⁴. One of the most recent studies⁴² in modelling cross-tissue networks highlighted coherent subnetworks that were not part of any of the single-tissue networks but, instead, specific to cross-tissue interactions, showing that modelling molecular interactions operating between tissues is critical if we hope to understand physiological states associated with disease.

Whereas classic molecular biology provided very narrow views connecting molecular entities to disease, today's technologies allow the generation of comprehensive snapshots of living systems, which in turn allows a more systems-level view of the molecular states underlying physiological

states associated with disease. In single experiments, we can now generate terabytes of genotype, sequence, gene expression, physiological and imaging data. The degree to which any one of these different data types informs our view of disease may vary, but these data types provide complementary views that are useful individually and potentially exceptionally valuable when considered collectively.

Disease-associated networks such as the MEMN comprise hundreds of genes interacting in complex ways that collectively associate with physiological states such as fat mass, insulin levels and atherosclerotic-lesion size. Such networks may be indicated to cause variations in disease-associated traits and can also respond to (or sense) genetic and environmental variations that influence disease risk. For example, the MEMN was demonstrated to respond to a wide range of DNA variations in genes distributed throughout the genome and also responded to environmental perturbations such as changes in diet. For mice placed on a high-fat diet, more than 40% of the RNA traits that changed relative to those of mice on a normal, chow diet were concentrated in the MEMN (the probability of this overlap occurring by chance was computed to be $<10^{-200}$).

Perspectives

The disease-associated molecular networks that we can construct today are necessarily based on grossly incomplete sets of data. Even given the ability to assay DNA and RNA variation in whole populations in a comprehensive manner, the information is not complete, because we are far from completely characterizing rare variation, DNA variation other than SNP and copy number, variation in non-coding RNA levels and variation in the different isoforms of genes in any sample, much less in entire populations. Beyond DNA and RNA, it is not possible with existing technologies to measure all protein-associated traits or all the interactions between proteins and DNA/RNA, metabolite levels and other molecular entities important to the functioning of living systems. Furthermore, the types of high-dimensional data we are able to generate routinely today in populations represent only a snapshot at a single time point, which may allow the identification of the functional units of the system under study and how these units relate to one another but does not allow a complete understanding of how the functional units are put together or the mechanistic underpinnings of the complex set of functions carried out by individual cells, by entire organs and by whole systems comprising multiple organs.

Technological advances, however, allow the generation of increasingly higher dimensional data, so we continue to progress towards a

more complete understanding of human disease. The next-generation sequencing technologies are already having a major impact on DNA sequencing, identifying rare variations in tumour tissues associated with different cancer types^{31,32}. In addition, subsequent generations of sequencing technologies are on the horizon and promise to deliver the sequence of entire human genomes in days and at a reasonable cost⁴⁷. Sequencing technologies can also be used to identify patterns of methylation⁴⁸, to fully characterize the transcriptome⁴⁹ and to identify transcripts that are being actively translated⁵⁰. The advances of the sequencing revolution therefore stand ready to provide unprecedented snapshots of complex systems that will allow a more accurate network view, which in turn will lead to models of disease that have greater predictive power.

One area in need of development regarding network-based approaches centres on the interpretation of high-dimensional data from which complex relationships and mathematical models are derived. The genomics field generally has been plagued by examples in which high-dimensional data have resulted in an unacceptably high rate of false positives. One striking example of this is a study that was undertaken to replicate published associations between 85 DNA variants and acute coronary syndromes. Of the 85 variants tested, only 1 gave rise to a nominally significant *P* value, highlighting a complete lack of support for the hypothesis that any of the variants previously reported in scores of publications as associating with acute coronary syndromes truly did so⁵¹. This problem is exacerbated when linking genotypes scored on hundreds of thousands of markers with tens of thousands of molecular phenotypes. Furthermore, understanding how to validate the accuracy of network models, how to compare networks across multiple conditions, species and methods, and, importantly, how to enable researchers to benefit from these models, which they may not fully understand, are among the most pressing problems to address if we are to move forwards. These issues are beginning to be addressed⁴⁴, and efforts such as the Dialogue for Reverse Engineering Assessments and Methods are making rapid progress in catalysing the type of interaction needed between experiment and theory to assess the accuracy of biological networks⁵².

Ultimately, our ability to construct predictive disease models will depend on our mastering the large-scale information being collected on systems relevant to disease. To accomplish this, data sharing must be more open, not only within industry but also within academic communities, where strong incentives to restrict data distribution exist to maintain competitive advantages. In addition, the development of tools and software platforms that allow the integration of large-scale, diverse data sets into complex models that can then be operated upon and refined by experimentalists in an iterative fashion is perhaps the most critical milestone we must reach in the biological sciences if large-scale data and results are to impact on biological research routinely at all levels.

The primary aims of generating and mining large-scale biological data sets are to learn the fundamental rules that govern complex living systems and to derive, as a result, predictive models of their behaviour. Without sophisticated mathematical algorithms capable of appropriately integrating the large-scale data, and without high-performance computing environments in which to apply these algorithms, it will be difficult to build generally predictive models. Information-systems support services will become increasingly critical both for building predictive models and for representing complex states of knowledge and making such knowledge accessible to researchers so that they may refine and correct the models of disease. Recent successes in programming machines to mine complex data to derive the fundamental laws of motion⁵³ perhaps represent a glimpse into the future of biology, in which machines may be able to derive fundamental rules in complex living systems, given large-scale data sets. The complexity of disease mechanisms must be recognized with investments in research directed towards these types of approach, which take a more holistic view in identifying the molecular networks that underlie physiological states associated with disease. Although systems approaches are still in their infancy, as a matter of necessity they will be viewed more and more as a crucial step towards an understanding of complex biological processes such as disease. ■

1. McKusick, V. A. *Mendelian Inheritance in Man: A Catalog of Human Genes and Genetic Disorders* (Johns Hopkins Univ. Press, 1998).
2. Chen, Y. *et al.* Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429–435 (2008).
This paper was the first demonstration that coherent networks of genes respond to genetic and environmental perturbations and in turn influence disease-associated traits, directly showing that common forms of disease are probably emergent properties of networks rather than the result of single gene changes.
3. Altshuler, D., Daly, M. J. & Lander, E. S. Genetic mapping in human disease. *Science* **322**, 881–888 (2008).
4. Barrett, J. C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nature Genet.* **40**, 955–962 (2008).
5. Zeggini, E. *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature Genet.* **40**, 638–645 (2008).
6. Kathiresan, S. *et al.* Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nature Genet.* **40**, 189–197 (2008).
7. Willer, C. J. *et al.* Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nature Genet.* **40**, 161–169 (2008).
8. Haiman, C. A. *et al.* A common genetic risk factor for colorectal and prostate cancer. *Nature Genet.* **39**, 954–956 (2007).
9. Haiman, C. A. *et al.* Multiple regions within 8q24 independently affect risk for prostate cancer. *Nature Genet.* **39**, 638–644 (2007).
10. Li, M. *et al.* CFH haplotypes without the Y402H coding variant show strong association with susceptibility to age-related macular degeneration. *Nature Genet.* **38**, 1049–1054 (2006).
11. Maller, J. *et al.* Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nature Genet.* **38**, 1055–1059 (2006).
12. Thorleifsson, G. *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nature Genet.* **41**, 18–24 (2009).
13. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
This paper is a confirmation in a human population that common diseases like obesity are the result of complex molecular networks responding to genetic and environmental perturbations.
14. Schadt, E. E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107 (2008).
15. Lum, P. Y., Derry, J. M. & Schadt, E. E. Integrative genomics and drug development. *Pharmacogenomics* **10**, 203–212 (2009).
16. Schadt, E. E., Friend, S. H. & Shaywitz, D. A. A network view of disease and compound screening. *Nature Rev. Drug Discov.* **8**, 286–295 (2009).
17. Goldstein, D. B. Common genetic variation and human traits. *N. Engl. J. Med.* **360**, 1696–1698 (2009).
18. Hardy, J. & Singleton, A. Genomewide association studies and human disease. *N. Engl. J. Med.* **360**, 1759–1768 (2009).
19. Kraft, P. & Hunter, D. J. Genetic risk prediction — are we there yet? *N. Engl. J. Med.* **360**, 1701–1703 (2009).
20. Moffatt, M. F. *et al.* Genetic variants regulating *ORMDL3* expression contribute to the risk of childhood asthma. *Nature* **448**, 470–473 (2007).
This was among the first studies to identify a disease-susceptibility gene by restricting attention to DNA variants that simultaneously associate with the disease and the expression levels of genes in the neighbourhood of the disease-associated variant.
21. Schadt, E. E. *et al.* An integrative genomics approach to infer causal associations between gene expression and disease. *Nature Genet.* **37**, 710–717 (2005).
This was the first study to demonstrate that causal relationships between molecular-profiling traits (such as gene expression) and disease traits could be systematically inferred by integrating these data with genotypic data in human and experimental populations.
22. Schadt, E. E. *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297–302 (2003).
23. Monks, S. A. *et al.* Genetic inheritance of gene expression in human cell lines. *Am. J. Hum. Genet.* **75**, 1094–1105 (2004).
24. Morley, M. *et al.* Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**, 743–747 (2004).
25. Foss, E. J. *et al.* Genetic basis of proteome variation in yeast. *Nature Genet.* **39**, 1369–1375 (2007).
26. Fraser, H. B. & Xie, X. Common polymorphic transcript variation in human disease. *Genome Res.* **19**, 567–575 (2009).
27. Smirnov, D. A., Morley, M., Shin, E., Spielman, R. S. & Cheung, V. G. Genetic analysis of radiation-induced changes in human gene expression. *Nature* **459**, 587–591 (2009).
28. Mehrabian, M. *et al.* Integrating genotypic and expression data in a segregating mouse population to identify 5-lipoxygenase as a susceptibility gene for obesity and bone traits. *Nature Genet.* **37**, 1224–1233 (2005).
29. Yang, X. *et al.* Validation of candidate causal genes for abdominal obesity that affect shared metabolic pathways and networks. *Nature Genet.* **41**, 415–423 (2009).
30. Goldstein, D. B. Genomics and biology come together to fight HIV. *PLoS Biol.* **6**, e76 (2008).
31. Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**, 1801–1806 (2008).
32. Parsons, D. W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807–1812 (2008).
33. Zhu, J. *et al.* Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nature Genet.* **40**, 854–861 (2008).
This paper generalizes the early idea of integrating gene expression and DNA-variation data to infer causal relationships among gene expression traits and between gene

- expression and disease traits by integrating diverse types of data, including genotype, gene expression, protein-interaction and DNA-protein-binding data.
34. Keller, M. P. *et al.* A gene expression network model of type 2 diabetes links cell cycle regulation in islets with diabetes susceptibility. *Genome Res.* **18**, 706–716 (2008).
 35. Meng, H. *et al.* Identification of *Abcc6* as the major causal gene for dystrophic cardiac calcification in mice through integrative genomics. *Proc. Natl Acad. Sci. USA* **104**, 4530–4535 (2007).
 36. Ghazalpour, A. *et al.* Genomic analysis of metabolic pathway gene expression in mice. *Genome Biol.* **6**, R59 (2005).
 37. Ghazalpour, A. *et al.* Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet.* **2**, e130 (2006).
 38. Zhu, J. *et al.* An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenet. Genome Res.* **105**, 363–374 (2004).
 39. Rual, J. F. *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005).
 40. Han, J. D. *et al.* Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **430**, 88–93 (2004).
 41. Gargalovic, P. S. *et al.* Identification of inflammatory gene modules based on variations of human endothelial cell responses to oxidized lipids. *Proc. Natl Acad. Sci. USA* **103**, 12741–12746 (2006).
 42. Dobrin, R. *et al.* Multi-tissue coexpression networks reveal unexpected subnetworks associated with disease. *Genome Biol.* **10**, R55 (2009).
 43. Pe'er, D., Regev, A., Elidan, G. & Friedman, N. Inferring subnetworks from perturbed expression profiles. *Bioinformatics* **17** (suppl. 1), S215–S224 (2001).
 44. Zhu, J. *et al.* Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations. *PLOS Comput. Biol.* **3**, e69 (2007).
 45. Schadt, E. E., Sachs, A. & Friend, S. Embracing complexity, inching closer to reality. *Sci. STKE* **2005**, pe40 (2005).
 46. Zeyda, M. & Stulnig, T. M. Adipose tissue macrophages. *Immunol. Lett.* **112**, 61–67 (2007).
 47. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009).
 48. Cokus, S. J. *et al.* Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**, 215–219 (2008).
 49. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Rev. Genet.* **10**, 57–63 (2009).
 50. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
 51. Morgan, T. M., Krumholz, H. M., Lifton, R. P. & Spertus, J. A. Nonvalidation of reported genetic risk factors for acute coronary syndrome in a large-scale replication study. *J. Am. Med. Assoc.* **297**, 1551–1561 (2007).
 52. Stolovitsky, G. & Califano, A. (eds). *Reverse Engineering Biological Networks: Opportunities and Challenges in Computational Methods for Pathway Inference* (Wiley, 2007).
 53. Schmidt, M. & Lipson, H. Distilling free-form natural laws from experimental data. *Science* **324**, 81–85 (2009).
 54. Bock, G. & Goode, J. A. (eds). 'In Silico' *Simulation of Biological Processes* 91–103; 119–128; 244–252 (Wiley, 2002).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence should be addressed to E.E.S. (eschadt@pacificbiosciences.com).

Co-translational mRNA decay in *Saccharomyces cerevisiae*

Wenqian Hu^{1*}, Thomas J. Sweet^{1*}, Sangpen Chamnongpol², Kristian E. Baker¹ & Jeff Collier¹

The rates of RNA decay and transcription determine the steady-state levels of all messenger RNA and both can be subject to regulation. Although the details of transcriptional regulation are becoming increasingly understood, the mechanism(s) controlling mRNA decay remain unclear. In yeast, a major pathway of mRNA decay begins with deadenylation followed by decapping and 5'–3' exonuclease digestion. Importantly, it is hypothesized that ribosomes must be removed from mRNA before transcripts are destroyed. Contrary to this prediction, here we show that decay takes place while mRNAs are associated with actively translating ribosomes. The data indicate that dissociation of ribosomes from mRNA is not a prerequisite for decay and we suggest that the 5'–3' polarity of mRNA degradation has evolved to ensure that the last translocating ribosome can complete translation.

In eukaryotic cells, mRNA is predominately degraded by two alternative pathways that are both initiated by shortening of the 3' polyadenosine tail (deadenylation). After deadenylation, either the 5' 7mGpppN cap is removed (decapping) and the message is digested exonucleolytically 5'–3' or the transcript is destroyed 3'–5' by the cytoplasmic exosome¹. The two mechanisms of mRNA decay together determine basal mRNA levels, thereby significantly contributing to overall gene expression.

Translation is postulated to be a key determinant in controlling mRNA decapping¹. The translational initiation complex eIF-4F occupies the cap during translation, which suggests that its binding must be antagonized and translational repression must ensue before decapping can occur^{1–4}. This hypothesis is supported by several observations. First, translational initiation rate is inversely proportional to decapping rate³. Second, the decapping regulators Dhh1p and Pat1p are translational repressors and their role in promoting mRNA decapping is partly a function of this activity^{5,6}. Third, mRNA decapping can occur at an unquantified level in ribosome-free cellular foci, termed P-bodies². Collectively, a two-step model for mRNA decay has been proposed where ribosome dissociation is a necessary first step before mRNA decapping^{1–4}.

Deadenylated mRNA remains on polyribosomes

The aforementioned model for mRNA decay predicts that after deadenylation but before decapping a ribosome-free state exists^{1–4}. We reasoned that in a decapping-defective cell (*dcp2Δ*), deadenylated RNA would accumulate in this ribosome-free state. We used sucrose density gradients to survey mRNA ribosome association in wild-type and decapping-defective cells (*dcp2Δ*). Greater than 90% of total cellular mRNA is analysed by this method (data not shown), and ribosome-free ribonucleoprotein (RNP) structures can be clearly separated from polyribosomes (Supplementary Fig. 2c). As predicted, inhibition of decapping did result in accumulation of deadenylated mRNA (Supplementary Fig. 2a, b, f); however, the mRNAs continued to sediment deep into a sucrose gradient even when deadenylated (Supplementary Fig. 2d, g, h). In fact, the sedimentation profiles of several mRNAs in *dcp2Δ* cells were indistinguishable from those in wild-type cells (Supplementary Fig. 2d, g, h). The rapid

sedimentation of these RNAs could occur either because they were sequestered in heavy particles (perhaps P-bodies)^{1,2} or because they were associated with ribosomes. The fact that sedimentation correlated with the length of the open reading frame (ORF) (Supplementary Fig. 2d, g, h) strongly suggested that the mRNAs were ribosome associated (see below).

Decapped mRNAs are found on polyribosomes

Because deadenylated mRNAs are the substrates for decapping³ we also assessed the sedimentation profiles of decapped RNAs. This was done in cells defective for the 5'–3' exonuclease (*xrn1Δ*). In these cells a stable decapped decay intermediate shortened by two nucleotides accumulates (indicated by '–cap'; Fig. 1a) and can be detected by using quantitative primer extension analysis (Supplementary Fig. 10)^{7–9}. Interestingly, the decapped intermediate showed the same sedimentation profile as the deadenylated RNA (Fig. 1a versus Supplementary Fig. 2); most (83–95%) decapped mRNA being present in polyribosomes (Fig. 1a, d). To determine whether the decay intermediate was associated with ribosomes, we took four approaches. First, introduction of a premature termination codon that shortened the ORF of *PGK1* by 393 codons resulted in a dramatic shift to significantly lighter fractions (Fig. 1b, c). Second, introduction of a stem-loop to limit translation⁵ caused a shift towards the top of the gradient both for capped and uncapped mRNAs (Fig. 1c). Third, treatment with EDTA (known to dissociate ribosomes) shifted the sedimentation to the top of the gradient (Fig. 1c). Finally, we showed that decapped mRNAs were associated with ribosomes by ribosome immunoprecipitation¹⁰ (Supplementary Fig. 3).

To investigate ribosome-associated decapping further and to exclude the possibility that decapping had occurred before initiation of protein synthesis, we took a transcriptional-pulse chase approach using the *PGK1* mRNA reporter⁹. Using a circularization-based PCR with reverse transcription (cRT–PCR)¹¹ analysis, we noted that decapped RNA started to appear around 60 min after initiation of transcription (Fig. 2a–c). Separation of cell lysate into non-translating and polyribosome-associated fractions indicated that when decapping is initiated at 60 min, most decapped mRNA was polyribosome associated (Fig. 2d). To exclude further the possibility that association of

¹Center for RNA Molecular Biology, Case Western Reserve University, Cleveland, Ohio 44106, USA. ²Affymetrix Inc., 26111 Miles Road, Cleveland, Ohio 44128, USA.

*These authors contributed equally to this work.

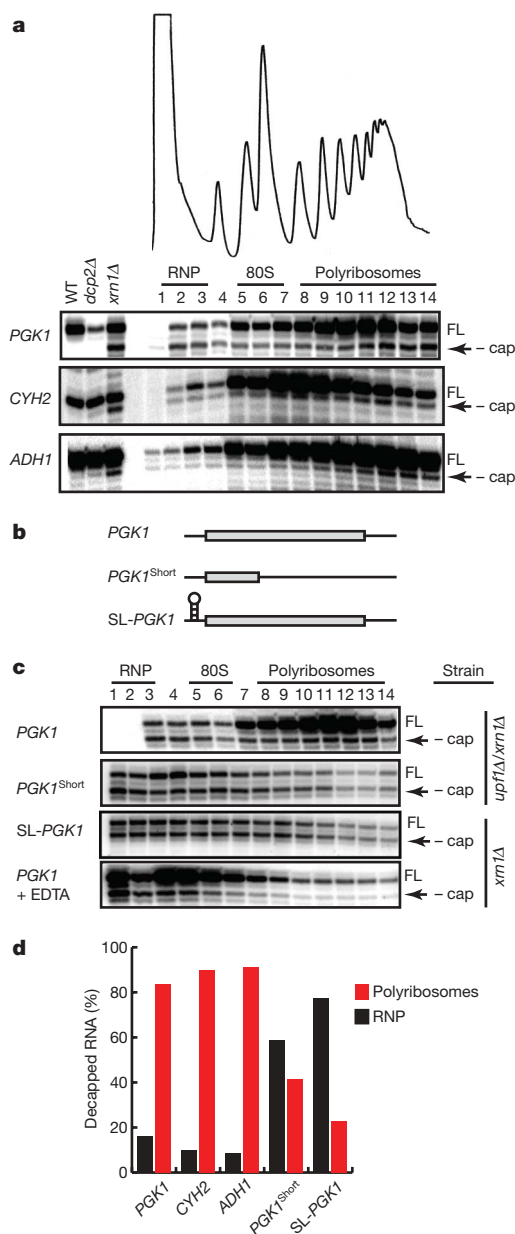


Figure 1 | Decapped mRNA is associated with polyribosomes. **a**, Primer extension analysis on endogenous *PGK1*, *CYH2* and *ADH1* mRNA was performed on RNA isolated from sucrose gradient fractions of an *xrn1Δ* cell lysate. RNP, 80S and polyribosomes are indicated above fraction numbers. FL, full-length mRNA; '–cap', decapped mRNA. Primer extension analyses on total RNA (15 μg) from wild-type, *dcp2Δ* and *xrn1Δ* cells are shown on the left side of each panel to indicate –cap mRNA is observed only in *xrn1Δ* cells. **b**, Representation of *PGK1* reporter, *PGK1* reporter with a PTC (*PGK1*^{Short}), and *PGK1* reporter with a stem-loop in its 5' untranslated region (SL-*PGK1*). **c**, Primer extension on RNA from sucrose gradient fractions from lysates of *upf1Δ/xrn1Δ* cells expressing *PGK1* reporter or *PGK1*^{Short} reporter, and from *xrn1Δ* cells expressing SL-*PGK1* or *PGK1* reporter. In the bottom panel, lysates from *xrn1Δ* cells expressing the *PGK1* reporter were incubated in presence of 50 mM EDTA before loading on sucrose gradients. **d**, Quantification of –cap mRNAs as a percentage of total reverse transcription product in RNP and polyribosome fractions.

uncapped mRNA with polyribosomes is a consequence of reloading ribosomes, we used a transcriptional shut-off approach³ with the *PGK1* reporter and monitored decapping using primer extension analysis. Transcription was arrested and further translation was blocked by addition of cycloheximide. Because cycloheximide inhibits ribosome elongation, newly initiated strands would be arrested at 80S (ref. 12). Strikingly, mRNAs trapped on ribosomes continued to be

decapped until greater than 50% was achieved after 120 min (Fig. 2e, f). In the absence of cycloheximide, the bolus of newly decapped mRNA sediments to the top of the gradient by 120 min (Fig. 2g), indicating that ribosomal run-off ensued. These results clearly show that decapping can occur when mRNAs are associated with actively translocating ribosomes.

Wild-type decapping is co-translational

The foregoing studies were all conducted in *xrn1Δ* cells to allow for the enrichment of decay intermediates. To detect decay intermediates in wild-type cells, we designed a reporter with ten consecutive rare codons (*PGK1*^{RC}; Fig. 3a). We reasoned that the presence of rare codons might slow ribosome transit¹³ and result in accumulation of decapped, ribosome-associated decay intermediates¹⁴. Importantly, the *PGK1*^{RC} reporter's decay is dependent on decapping and is not a major substrate for No-Go mRNA decay (Supplementary Fig. 4)¹⁵. We analysed the *PGK1*^{RC} reporter on sucrose gradients and detected decay intermediates using high-resolution polyacrylamide gel electrophoresis (PAGE) followed by northern blot. Notably, using a 3' end-specific probe, decay intermediates of about 500 nucleotides were detected in the region of the gradient associated with a single ribosome (that is, 80S; Fig. 3b). In addition, mRNA intermediates of increasing length were also detected in polyribosome fractions and their size correlated well with possible ribosome occupancy (Fig. 3b). Addition of formaldehyde before cell lysis was used to ensure that the decay intermediates were generated *in vivo* (Fig. 3b); however, similar fragments were seen without formaldehyde treatment (Supplementary Fig. 6a). A probe complementary to the 5' end of the mRNA failed to detect decay intermediates, which confirmed that the truncated mRNA was trimmed from the 5' end (Fig. 3b and Supplementary Fig. 5). Most importantly, polyribosome-associated decay intermediates were lost in *dcp2Δ* and *xrn1Δ* mutants (Fig. 3c and Supplementary Fig. 6b), which indicates their formation requires mRNA decapping and 5'–3' exonucleolytic digestion. Moreover, the *PGK1*^{RC} mRNA decay fragments were not a result of No-Go decay¹⁵ (Supplementary Fig. 6c).

We used four experiments to demonstrate that the sedimentation pattern of the *PGK1*^{RC} mRNA decay intermediates is a result of polyribosome association. First, we inhibited translation of the mRNA. Insertion of a stem-loop structure into the 5' untranslated region (SL-*PGK1*^{RC}; Supplementary Fig. 7a) shifted the full-length mRNA to the top of the gradient, and no decay intermediates were detectable deep in the gradient (Supplementary Fig. 7b). Second, we terminated ribosome elongation before rare-codon recognition by introduction of a stop codon upstream of the rare codon stretch (*PGK1*^{PTC-RC}; Supplementary Fig. 7a). This experiment was performed in *upf1Δ* cells to prevent nonsense-mediated decay¹⁶. Terminating ribosome translocation before the rare codons completely inhibited the formation of polyribosome-associated decay intermediates (Supplementary Fig. 7c compared with d). Further demonstrating that ribosome recognition of the rare-codon stretch is required, repositioning the rare codon stretch within the *PGK1* ORF resulted in a predictable size shift in polyribosome-associated decay fragments (Supplementary Fig. 8). Finally, we performed affinity purification of polyribosomes¹⁰ and demonstrated that the decay fragments are ribosome bound (Supplementary Fig. 9). In sum, these data strongly demonstrate that decapping can be detected on polyribosomes in wild-type cells if translational elongation is slowed *in cis*.

Endogenous mRNAs are decapped on polysomes

The foregoing experiment used a reporter harbouring rare codons. To determine whether endogenous mRNAs in wild-type cells were also decapped when associated with ribosomes, we developed a splinted ligation assay followed by RT-PCR (Fig. 3d). The RNA ligation mediated by the DNA splint is sequence specific¹⁷, thereby allowing us directly to detect the transient product generated by the

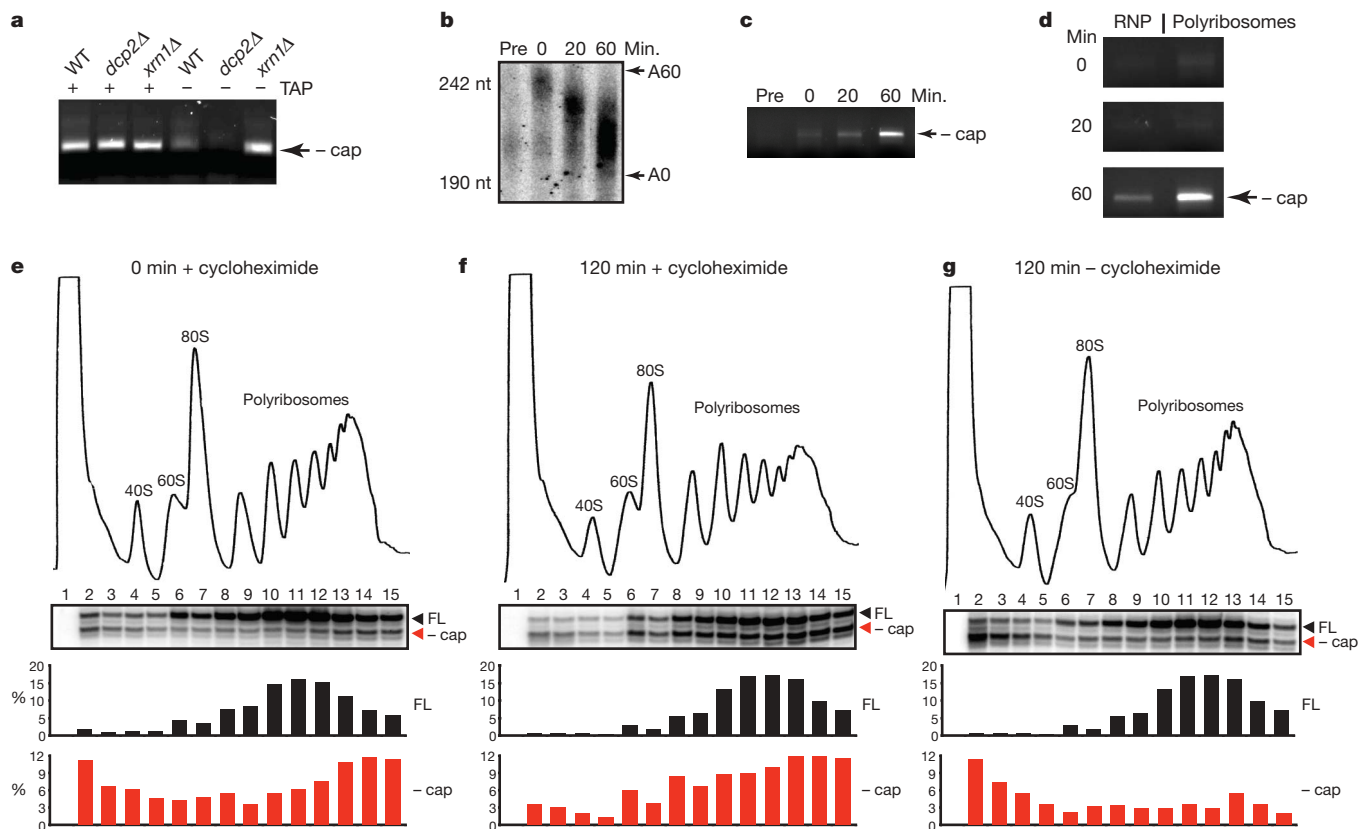


Figure 2 | mRNA decapping is initiated on polyribosomes. All experiments in Fig. 2 were performed using cells expressing *PGK1* reporter under control of the *GAL1* promoter. **a**, Total RNA from wild-type, *dcp2Δ*, and *xrn1Δ* cells was treated with (+) or without (–) tobacco acid pyrophosphatase (TAP), and cRT-PCR was performed to detect decapped *PGK1* reporter. **b–d**, Transcriptional pulse-chase of *PGK1* performed in *xrn1Δ* cells. **b**, Poly(A) tail status of *PGK1* was analysed by oligonucleotide-directed RNase H cleavage, PAGE and northern analysis. Pre indicates pre-induction. **c**, Decapping of *PGK1* mRNA monitored by cRT-PCR. **d**, Cell lysates from the pulse-chase were separated on sucrose gradients. RNA from

gradient fractions was pooled into non-translating (RNP) and polysome pools and decapped *PGK1* was detected by cRT-PCR. **e–g**, Transcriptional shut-off of *PGK1* was performed in *xrn1Δ* cells. Lysates from cells at 0 min after shut-off (**e**), 120 min after shut-off in the presence of 25 $\mu\text{g ml}^{-1}$ cycloheximide (**f**) and 120 min after shut-off without cycloheximide (**g**) were separated by sucrose gradients. RNA from gradient fractions was analysed by primer extension for *PGK1* reporter. The quantifications of full length (FL) and decapped (–cap) mRNA as a percentage of total extension product are shown for each time point.

decapping reaction (that is, an RNA with 5' phosphate). Using this assay, decapped products from endogenous *PGK1* and *RPL41A* mRNA were detected in wild-type cells (Fig. 3e). A product was not detected in *dcp2Δ* cells (Fig. 3e), which indicates that formation requires decapping *in vivo*. Consistent with this, *in vitro* removal of the 5' cap by tobacco acid pyrophosphatase resulted in detection of RT-PCR products both in wild-type and *dcp2Δ* cells (Fig. 3e). Together, these data indicate that the splinted ligation/RT-PCR assay monitors 5' decapping. We performed this assay on RNA recovered from sucrose gradient fractions of wild-type cell lysate, and found that the decapped mRNAs from endogenous *PGK1* and *RPL41A* were predominately detected on polyribosomes (Fig. 3f). Notably, the sedimentation pattern of the decapped mRNA correlates with the total mRNA detected by northern blot (Fig. 3f) and mRNA ORF length (Fig. 3f). Consistent with our earlier findings (Fig. 2), the sedimentation of decapped mRNA on polyribosomes is unlikely to be a result of ribosome reloading because the decapped intermediate is exceptionally transient in a wild-type cell. Collectively, these data indicate that in wild-type cells, endogenous mRNAs are decapped on polyribosomes.

Conclusions and perspective

In sum, we have shown that decapping and 5'–3' degradation of mRNA can occur when the transcripts are associated with actively translating ribosomes (Supplementary Fig. 1). Co-translational degradation of mRNA has been previously hypothesized^{18,19}. Here

we experimentally demonstrate this hypothesis and show mRNA remains associated with active ribosomes during the process of mRNA decapping and exonucleolytic degradation. The data clearly indicate that sequestration into a ribosome-free state (for example, P-bodies) is not a prerequisite for initiation of mRNA decay. These findings are consistent with the demonstration in yeast, *Drosophila* and humans that mRNA metabolism can be uncoupled from P-body formation^{6,20–22}. Moreover, they also help to explain why decay factors (for example hDCP2 and Xrn1p) have been found to co-sediment with polyribosomes^{19,23}. Our findings raise several interesting mechanistic questions, for instance how mRNA half-lives are determined in the context of ongoing translation. Moreover, it is unclear how the decapping machinery associates and functions on an actively translating mRNA. Interestingly, it has previously been proposed that decapping regulators promoted a ribosome-free state^{1,2,5,6}; it now seems likely that they function in response to as yet unknown cues to render the cap more accessible to the decapping enzyme during translation (Supplementary Fig. 1).

Finally, we note that co-translational mRNA degradation makes sense from an evolutionary point of view. Specifically, the three steps of decay each serve systematically to limit translational events without interfering with them. Deadenylation may reduce translational efficiency, perhaps through loss of the poly(A) binding protein, Pab1p¹⁶ or association of decapping regulators⁵. mRNA decapping inhibits further translation initiation events. Finally, degradation from the 5' end while the mRNA is ribosome associated ensures

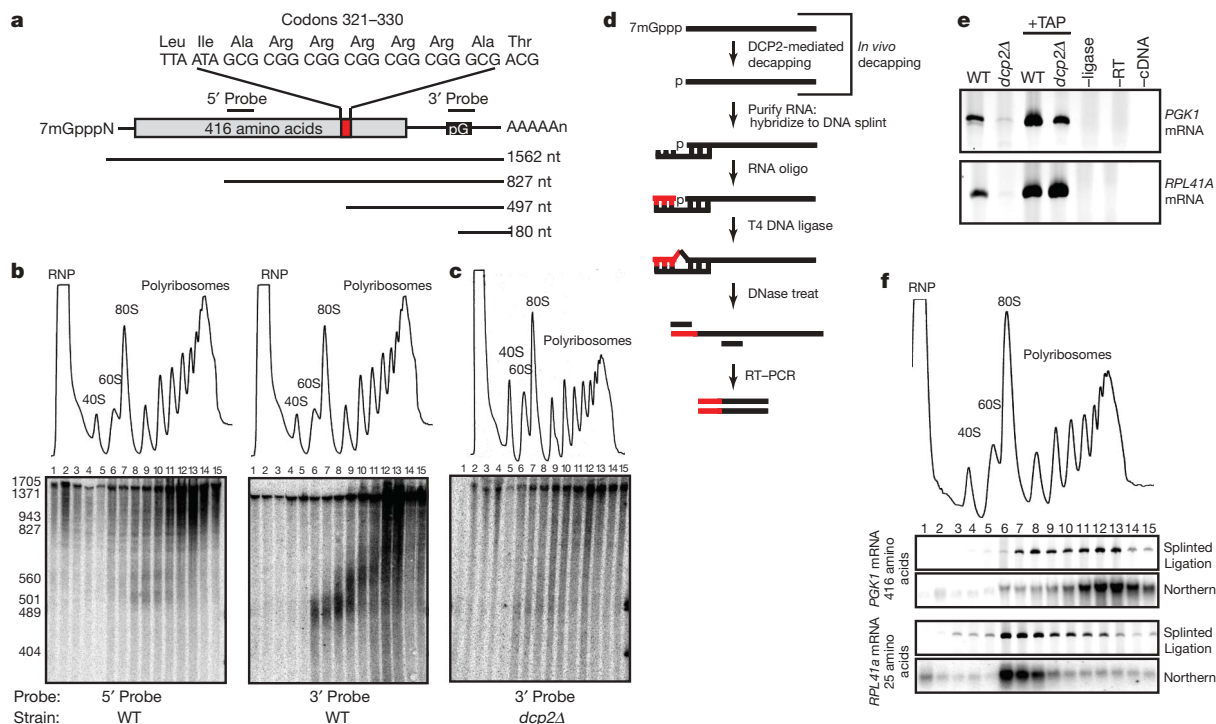


Figure 3 | mRNA decapping occurs on polyribosomes in wild-type cells. **a**, The *PGK1^{RC}* reporter is depicted. **b**, Northern blot analysis of *PGK1^{RC}* mRNA after sucrose gradient fractionation. RNA detected using a 5' or 3' probe as depicted in **a**. WT, wildtype. **c**, The same analysis as in **b** performed in *dcp2Δ* cells. **d**, Splinted-ligation RT-PCR assay to detect endogenous decapped mRNA in wild-type cells. An RNA adaptor is ligated specifically to decapped mRNA by a DNA splint by T4 DNA ligase. The DNA splint is removed by DNase I treatment and the ligation product is detected by

RT-PCR using a gene and adaptor-specific primers. The PCR product is indicative of decapped mRNA. **e**, Splinted-ligation RT-PCR analysis for endogenous *PGK1* and *RPL41A* mRNAs on total RNA from wild-type and *dcp2Δ* cells. +TAP, total RNA treated with tobacco acid pyrophosphatase to remove the 5' cap *in vitro*; –ligase, no T4 ligase; –RT, no reverse transcriptase; –cDNA, no cDNA template added to PCR. **f**, RNA recovered from sucrose gradient fractions of wild-type cell lysate was analysed by splinted-ligation RT-PCR or northern blot using gene-specific probes.

decay does not impede residual ribosomes undergoing translocation. In this way, the final polypeptide expressed before the mRNA is destroyed is full length and functional.

METHODS SUMMARY

All experiments were performed using early log phase cells grown at 24 °C in synthetic medium containing appropriate sugars. RNA and polysome analyses were performed as described previously⁵. The cRT-PCR assay was performed as described previously¹¹ with oJC620 for reverse transcription and oJC620/oJC635 for PCR amplification. The *PGK1^{RC}* reporter was generated from fragments amplified from a previously described *PGK1* reporter⁹ using oJC558/oJC556 and oJC557/oJC559; fragments were combined to produce a template for amplification of full-length *PGK1^{RC}* using oJC558/oJC559, followed by cloning onto the *PGK1* reporter backbone at the BamHI and HindIII sites. Affinity purification of polyribosomes was performed as described previously¹⁰. Detection of endogenous decapped mRNA was achieved by ligating an RNA adaptor (oJC706) to the 5' end of decapped mRNA by splinted ligation, removing the DNA splint by DNase I, complementary DNA (cDNA) synthesis by Superscript II reverse transcriptase using a gene-specific primer, and DNA amplification by PCR using a primer complementary to the RNA adaptor and a gene-specific primer.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 June; accepted 6 July 2009.

Published online 23 August 2009.

1. Franks, T. M. & Lykke-Andersen, J. The control of mRNA decapping and P-body formation. *Mol. Cell* **32**, 605–615 (2008).
2. Parker, R. & Sheth, U. P bodies and the control of mRNA translation and degradation. *Mol. Cell* **25**, 635–646 (2007).
3. Collier, J. & Parker, R. Eukaryotic mRNA decapping. *Annu. Rev. Biochem.* **73**, 861–890 (2004).

4. Eulalia, A., Behm-Ansmant, I. & Izaurralde, E. P bodies: at the crossroads of post-transcriptional pathways. *Nature Rev. Mol. Cell Biol.* **8**, 9–22 (2007).
5. Collier, J. & Parker, R. General translational repression by activators of mRNA decapping. *Cell* **122**, 875–886 (2005).
6. Chu, C. Y. & Rana, T. M. Translation Repression in Human Cells by MicroRNA-Induced Gene Silencing Requires RCK/p54. *PLoS Biol.* **4**, e210 (2006).
7. He, F. & Jacobson, A. Upf1p, Nmd2p, and Upf3p regulate the decapping and exonucleolytic degradation of both nonsense-containing mRNAs and wild-type mRNAs. *Mol. Cell Biol.* **21**, 1515–1530 (2001).
8. Hsu, C. L. & Stevens, A. Yeast cells lacking 5'→3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. *Mol. Cell Biol.* **13**, 4826–4835 (1993).
9. Muhrad, D., Decker, C. J. & Parker, R. Turnover mechanisms of the stable yeast *PGK1* mRNA. *Mol. Cell Biol.* **15**, 2145–2156 (1995).
10. Inada, T. *et al.* One-step affinity purification of the yeast ribosome and its associated proteins and mRNAs. *RNA* **8**, 948–958 (2002).
11. Couttet, P., Fromont-Racine, M., Steel, D., Pictet, R. & Grange, T. Messenger RNA deadenylation precedes decapping in mammalian cells. *Proc. Natl Acad. Sci. USA* **94**, 5628–5633 (1997).
12. Amrani, N., Ghosh, S., Mangus, D. A. & Jacobson, A. Translation factors promote the formation of two states of the closed-loop mRNP. *Nature* **453**, 1276–1280 (2008).
13. Smith, D. W. & McNamara, A. L. Specialization of rabbit reticulocyte transfer RNA content for hemoglobin synthesis. *Science* **171**, 577–579 (1971).
14. Stevens, A. 5'-exoribonuclease 1: Xrn1. *Methods Enzymol.* **342**, 251–259 (2001).
15. Doma, M. K. & Parker, R. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* **440**, 561–564 (2006).
16. Jacobson, A. & Peltz, S. W. Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells. *Annu. Rev. Biochem.* **65**, 693–739 (1996).
17. Moore, M. J. & Query, C. C. Joining of RNAs by splinted ligation. *Methods Enzymol.* **317**, 109–123 (2000).
18. Beelman, C. A. & Parker, R. Differential effects of translational inhibition in cis and in trans on the decay of the unstable yeast *MFA2* mRNA. *J. Biol. Chem.* **269**, 9687–9692 (1994).
19. Mangus, D. A. & Jacobson, A. Linking mRNA turnover and translation: assessing the polyribosomal association of mRNA decay factors and degradative intermediates. *Methods* **17**, 28–37 (1999).

20. Decker, C. J., Teixeira, D. & Parker, R. Edc3p and a glutamine/asparagine-rich domain of Lsm4p function in processing body assembly in *Saccharomyces cerevisiae*. *J. Cell Biol.* **179**, 437–449 (2007).
21. Sweet, T. J., Boyer, B., Hu, W., Baker, K. E. & Collier, J. Microtubule disruption stimulates P-body formation. *RNA* **13**, 493–502 (2007).
22. Eulalio, A., Behm-Ansmant, I., Schweizer, D. & Izaurralde, E. P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Mol. Cell. Biol.* **27**, 3970–3981 (2007).
23. Wang, Z., Jiao, X., Carr-Schmid, A. & Kiledjian, M. The hDcp2 protein is a mammalian mRNA decapping enzyme. *Proc. Natl Acad. Sci. USA* **99**, 12663–12668 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Maroney and T. Nilsen for sharing unpublished data about the poly(A) tailing assay. We also thank T. Nilsen for his insight, suggestions and evaluation of the manuscript. We also thank R. Parker, A. van Hoof and M. Wickens for reagents and advice. Funding was provided by the American Heart Association and the National Institutes of Health.

Author Contributions W.H., K.E.B. and J.C. wrote the manuscript. W.H. and T.J.S. performed the experiments. S.C. provided technical expertise with poly(A) tailing assays. All of the authors contributed to discussion and the design of the research. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.C. (jmc71@case.edu).

METHODS

Yeast strains and growth conditions. The genotypes of all yeast strains used in this study are listed in Supplementary Table 1. Unless indicated, all strains are based on BY4741. Cells were grown in standard synthetic medium (pH 6.5) supplemented with appropriate amino acids and either 2% glucose, 2% galactose/1% sucrose, or 2% sucrose as the carbon source. All cells were grown at 24 °C and collected at mid-log phase (3.0×10^7 cells ml⁻¹).

Plasmids and oligonucleotides. The plasmids and oligonucleotides used in this study are listed in Supplementary Table 2. To construct the *PGK1*^{RC} (pJC314) and *SL-PGK1*^{RC} (pJC320) reporters, DNA amplified from either pJC296 (*PGK1*) or pJC134 (*SL-PGK1*) using oligonucleotides oJC558/oJC556 and oJC557/oJC559 was combined and used as the template for amplification of full-length *PGK1* using oligonucleotides oJC558/oJC559. Full-length fragments were cloned into the HindIII and BamHI sites of pJC296. The *PGK1*^{RC390} (pJC372) was constructed in a similar manner using oJC558/oJC824 and oJC559/oJC825. The PTC was introduced into pJC314 by site-directed mutagenesis using oligonucleotides oJC611/oJC612, resulting in pJC327. The *PGK1*^{short} (pJC349) was made by introducing a stop codon into codon 21 of pJC331 using oJC676/oJC677.

Northern RNA analysis. Northern RNA analysis was performed as previously described⁵. For the mRNA half-life measurements in Supplementary Fig. 4A, cells with the *PGK1*^{RC} reporter (pJC314) were grown in 2% galactose, 1% sucrose synthetic media and collected at mid-log phase (3.0×10^7 cells ml⁻¹). Transcription repression was achieved by re-suspending collected cells in media containing 4% glucose. After transcriptional repression, cell aliquots were removed and isolated total RNA (30 µg) was analysed by electrophoresis through 1.4% formaldehyde agarose gel. For transcriptional pulse-chase experiments in Fig. 2b, yJC182 expressing pJC331 was grown in 2% sucrose synthetic media. At mid-log phase, 2% galactose was added and cells incubated for 10 min. Transcription was inhibited by collecting cells and re-suspending in 2% glucose-containing media. Aliquots were removed over time and isolated RNA (30 µg) analysed by 6% PAGE. In Fig. 3b, c and Supplementary Figs 5–9, yJC151 was transformed with the appropriate reporter plasmids and grown in synthetic media contain 2% galactose/1% sucrose. Cells were harvested at mid-log phase, RNA isolated and analysed (30 µg) by 6% PAGE. Northern analyses were performed using radiolabelled oligonucleotide or RNA probes. Specifically, endogenous *PGK1* was detected using oRP25 (ref. 9), endogenous *RPL41a* with oJC124, endogenous *MFA2* with an *in vitro* synthesized RNA from pJC313, *PGK1* reporters with oRP121, and U1snRNA using oJC652.

Polyribosome analysis. All polyribosome analysis (with the exception of those shown in Figs 2e–g and 3b) were performed as previously described⁵ but with the following modifications. Specifically, cells grown to mid-log phase were treated with cycloheximide to a final concentration of 100 µg ml⁻¹ and collected by centrifugation. Cell pellets were lysed in buffer (10 mM Tris, pH 7.4, 100 mM NaCl, 30 mM MgCl₂, 500 µg ml⁻¹ heparin, 1 mM DTT, 100 µg ml⁻¹ cycloheximide) by bead bashing, and Triton X-100 was added to a final concentration of 1%. All gradients were made on a Biocomp gradient maker and were 15–45% weight/weight (sucrose to buffer (50 mM TrisAcetate pH 7.0, 50 mM NH₄Cl, 12 mM MgCl₂, 1 mM DTT)). Unless otherwise indicated, 20 units (OD₂₆₀) of cell lysate were loaded onto each gradient. Gradients were centrifuged at 37,000 r.p.m. for 3 h at 4 °C in a Beckman SW-41Ti rotor and fractionated using a Brandel Fractionation System and an Isco UA-6 ultraviolet detector. RNA was isolated as previously described⁸. For cross-linking experiments (Fig. 3b), formaldehyde was added to mid-log phase cells to a final concentration of 0.25%. Cells

were incubated at 24 °C for 5 min (with shaking) before the addition of glycine to a final concentration of 125 mM to inhibit further cross-linking. Cells were further incubated at 24 °C for 10 min (with shaking) before collection. Cell lysis was performed as described (see above). For experiments in Fig. 2e–g, cells were grown in media containing 2% galactose/1% sucrose. At mid-log phase, cell growth media was exchanged with media containing 2% glucose to shut off transcription of the *PGK1* reporter. Where described, cycloheximide was added simultaneously to a final concentration of 25 µg ml⁻¹, culture aliquots removed, cells collected and RNA isolated and analysed as described above. For experiments in Fig. 2d, cells were grown to mid-log phase in media containing 2% sucrose, and reporter expression was induced by the addition of 2% galactose. Cells were incubated for 10 min before exchanging the media with that containing 2% glucose, and cell aliquots were removed at various times. Polyribosomes were isolated and analysed as described above, except that gradient fractions were pooled into RNP (fraction 1–5) and polyribosomes (fraction 6–14).

Ribosome and polyribosome affinity purification. Ribosomes and polyribosomes were affinity purified as previously described¹⁰. Using 500 µl anti-Flag agarose matrix (Sigma) and 20 units (OD₂₆₀) cell lysate.

Primer extension. Primer extensions were performed as previously described⁹. Briefly, polyribosome analysis was performed on cell lysate from *xrn1Δ* cells and RNA was extracted from each fraction. Primer extension was performed on 15 µg of total cellular RNA using SuperScript II (Invitrogen) and a radiolabelled oligonucleotide. Primer extension products were analysed on 8% polyacrylamide/7M urea gels followed by PhosphorImager analysis.

Poly(A) tailing assay. A schematic of the technique used to assay mRNA poly(A) tail length is shown in Supplementary Fig. 2e. Briefly, RNA purified from either unfractionated or fractionated whole-cell lysates was treated with purified yeast poly(A) polymerase and GTP:ITP to add a G:I tail to the 3' end of RNA. Reverse transcription (MMLV Reverse Transcriptase, USB) was performed with oJC639. Poly(A) tails were detected using oJC640 and a gene-specific forward primer for PCR (oJC791 for *MFA2* mRNA). The sample labelled A0 (unadenylated mRNA product) was generated by PCR using oJC789 and oJC790. PCR products were separated on 3% agarose gels followed by staining with SYBRGold (Invitrogen) or ethidium bromide. Stained gels were visualized and imaged using the ChemiGenius two-gel dock.

cRT-PCR. The cRT-PCR assay used to detect decapped mRNA has been described previously¹¹. Briefly, RNA purified from either unfractionated or fractionated whole-cell lysates was treated with T4 RNA Ligase (Promega). Ligated RNA was reverse transcribed through the junction by incubation with oJC620 and SuperScript II RT (Invitrogen) at 37 °C. cDNA was amplified by PCR using oJC620 and oJC635. PCR products were separated on 2% agarose gels followed by staining with ethidium bromide. Stained gels were visualized and imaged using the ChemiGenius two-gel dock.

Splint-ligation RT-PCR. Splinted-ligation of RNA was performed as described previously¹⁷. Briefly, an RNA adaptor (oJC706) was ligated to the free hydroxyl of a decapped mRNA facilitated by a gene-specific DNA splint. Ligation reactions were performed for 16 h at room temperature. After removal of the splint by DNase I treatment, the gene-specific primer was used for reverse transcription to synthesize cDNA using SuperScript II (Invitrogen). cDNA served as the template for PCR amplification using a primer complementary to the RNA adaptor (oJC707) and the gene-specific primer used in reverse transcription. PCR products were resolved by PAGE on 8% native gels and stained with ethidium bromide.

ARTICLES

An RNA-dependent RNA polymerase formed by TERT and the *RMRP* RNA

Yoshiko Maida¹, Mami Yasukawa¹, Miho Furuuchi¹, Timo Lassmann², Richard Possemato³, Naoko Okamoto¹,
Vivi Kasim¹, Yoshihide Hayashizaki², William C. Hahn^{3,4} & Kenkichi Masutomi^{1,5}

Constitutive expression of telomerase in human cells prevents the onset of senescence and crisis by maintaining telomere homeostasis. However, accumulating evidence suggests that the human telomerase reverse transcriptase catalytic subunit (TERT) contributes to cell physiology independently of its ability to elongate telomeres. Here we show that TERT interacts with the RNA component of mitochondrial RNA processing endoribonuclease (*RMRP*), a gene that is mutated in the inherited pleiotropic syndrome cartilage-hair hypoplasia. Human TERT and *RMRP* form a distinct ribonucleoprotein complex that has RNA-dependent RNA polymerase (RdRP) activity and produces double-stranded RNAs that can be processed into small interfering RNA in a Dicer (also known as DICER1)-dependent manner. These observations identify a mammalian RdRP composed of TERT in complex with *RMRP*.

Telomerase is a ribonucleoprotein complex that elongates telomeres. Although several proteins interact with telomerase¹⁻⁴, the minimal components of active telomerase include the catalytic telomerase reverse transcriptase (TERT) and a noncoding RNA (*TERC*) that encodes the template to synthesize telomeric DNA⁵. Telomere homeostasis mediated by telomerase maintains genomic stability and regulates cell lifespan⁶. Mutations in TERT, *TERC* or dyskerin, a telomerase-associated nucleolar protein involved in ribosomal RNA maturation⁷, are found in dyskeratosis congenita, a syndrome characterized by ectodermal dysplasia and bone marrow failure, and TERT mutations have been reported in aplastic anaemia and idiopathic pulmonary fibrosis⁸. Moreover, alterations in the regulation of telomeres and telomerase contribute to malignant transformation by affecting genomic integrity and cell immortalization⁶.

However, accumulating evidence suggests that TERT has activities beyond telomere maintenance^{9–13} and forms several intracellular complexes^{2–4}. In particular, the overexpression of TERT induces increased tumour susceptibility^{9,10} and disrupts stem-cell function independently of telomere maintenance¹², whereas the suppression of TERT expression alters global chromatin structure¹¹. Indeed, some of these telomere-independent functions of TERT do not require the expression of *TERC*¹².

Identification of a second RNA that interacts with TERT

To identify human TERT partners, we stably overexpressed a tandem affinity peptide (TAP)-tagged TERT protein in HeLa S3 cells, isolated TERT immune complexes, and identified a heterogeneous mixture of 38 RNA sequences associated with TERT (Supplementary Fig. 2 and Supplementary Table 1). We found that 5% of the sequences corresponded to *TERC* and the RNA component of mitochondrial RNA processing endoribonuclease (*RMRP*). *RMRP* is a 267-nucleotide noncoding RNA that is a small nucleolar RNA, like *TERC*, and is also found in mitochondria^{8,14}. *RMRP* mutations are found in the pleiotropic inherited syndrome, cartilage-hair hypoplasia¹⁵.

From a single immune complex, we confirmed that either over-expressed or endogenous TERT interacts with *RMRP* and *TERC*, by

isolating TAP-TERT (Fig. 1a) or endogenous TERT (Fig. 1b) complexes in both HeLa and 293T cells under conditions in which we failed to recover the ribozyme *RNase P*. We also found that the abundance of TERT-*RMRP* and TERT-*TERC* complexes was similar, even though *TERC* was expressed at five-fold higher levels than *RMRP* in these cells (Fig. 1c and Supplementary Fig. 3).

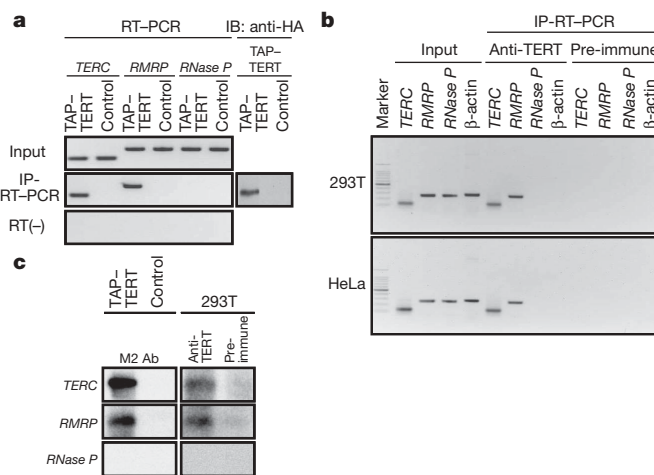


Figure 1 | TERT and RMRP interact. **a**, Detection of *RMRP* and *TERC*. RNA species associated with TAP–TERT complexes from a single immunoprecipitation (IP) were isolated and subjected to PCR with reverse transcription (RT–PCR). RT (–) indicates the absence of reverse transcriptase. Right panel shows the levels of TAP–TERT. HA, haemagglutinin; IB, immunoblot. **b**, TERT interacts with endogenous *RMRP*. TERT complexes from 293T and HeLa cells were isolated with an anti-TERT antibody and associated RNAs were subjected to RT–PCR. **c**, RNAs purified from TERT complexes isolated from HeLa S3 cells expressing TAP–TERT or a control vector or 293T cells were subjected to northern blotting. Ab, antibody.

¹Cancer Stem Cell Project, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. ²RIKEN Omics Science Center, RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan. ³Department of Medical Oncology, Dana-Farber Cancer Institute and Departments of Medicine, Brigham and Women's Hospital and Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA. ⁴Broad Institute of Harvard and MIT, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. ⁵PREST, Japan Science and Technology Agency, 4-1-8 Honcho Kawaguchi, Saitama 332-0012, Japan.

To characterize the interaction between TERT and *RMRP*, we used TERT truncation mutants and found that the amino terminal end of TERT (1–531) was necessary for interactions with *RMRP* (Supplementary Fig. 4). This region overlaps with two regions required for the binding of *TERC*^{8,16}. These observations demonstrate that TERT and *RMRP* form a new ribonucleoprotein complex distinct from the TERT–*TERC* enzyme.

The TERT–*RMRP* complex has RdRP activity

To test whether *RMRP* substitutes for *TERC* to reconstitute telomerase activity, we combined recombinant TERT with *TERC* or *RMRP* RNAs transcribed *in vitro*. Although we detected telomerase activity with TERT and *TERC* (Supplementary Fig. 5), we failed to detect telomerase activity when TERT and *RMRP* were co-incubated.

TERT has also been shown to act as a terminal transferase¹⁷, and human TERT shares sequence similarity to both viral reverse transcriptases and RdRPs¹⁸. RdRPs participate in the endogenous RNA interference (RNAi) pathway and in the regulation of post-transcriptional gene silencing^{19–23}. To examine whether the TERT–*RMRP* complex has RdRP and/or terminal transferase activity, we established an RNA synthesis activity assay with recombinant TERT protein (Supplementary Fig. 6) and RNA molecules transcribed *in vitro*. We predicted three modes that the TERT–*RMRP* complex might use to elongate RNA: (1) as an RdRP that uses a *de-novo*-synthesized RNA primer to elongate a complementary strand (Fig. 2a, left panel); (2) as an RdRP that uses a 3' fold-back (back-priming) configuration of template RNA as a primer (Fig. 2a, middle panel); or (3) as a terminal transferase (Fig. 2a, right panel). Viral RdRPs^{24,25} have been shown to use the first two modes to prime RdRP activity, and cellular RdRPs in

fission yeast²⁶ and fungi²³ use similar priming mechanisms to produce double-stranded (ds) RNAs that act as precursors for RNAi.

We found that recombinant TERT and *RMRP* produced two different products depending on the salt concentration (Fig. 2b and Supplementary Fig. 7). Specifically, we found ~267-nucleotide- (corresponding to sense *RMRP*) and ~534-nucleotide-sized products (hereafter referred to as sense plus antisense *RMRP* products) under high salt conditions, and *RMRP*-sized products under low salt conditions. To discriminate between these modes, we treated the products of the RdRP assay with RNase T1 (Fig. 2c) using conditions that favour the digestion of single-stranded RNA. RNase T1 treatment eliminated the ~267-nucleotide *RMRP*-sized RNA products produced under low salt concentrations (data not shown), indicating that [³²P]UTP was incorporated by terminal transferase activity.

In contrast, under high salt conditions, we found two RNAs (~267 and ~534 nucleotides) that collapsed into a single ~267-nucleotide band after treatment with RNase T1 (Fig. 2c). To eliminate the possibility that the sense plus antisense product represented partially denatured RNAs, we treated the products of the RdRP assay with bacterial RNase III to digest dsRNA, and found that only the input ~267-nucleotide RNA remained (Fig. 2d). Furthermore, when we left out adenine or guanine ribonucleotides, we failed to detect the sense plus antisense product (Fig. 2e). These observations confirm that the ~534-nucleotide sense plus antisense products are formed by RdRP activity and represent a double-stranded hairpin structure created by an RNA molecule composed of sense and antisense strands of *RMRP*.

To confirm that the interaction between TERT and *RMRP* was required for RdRP activity, we performed an RdRP activity assay using

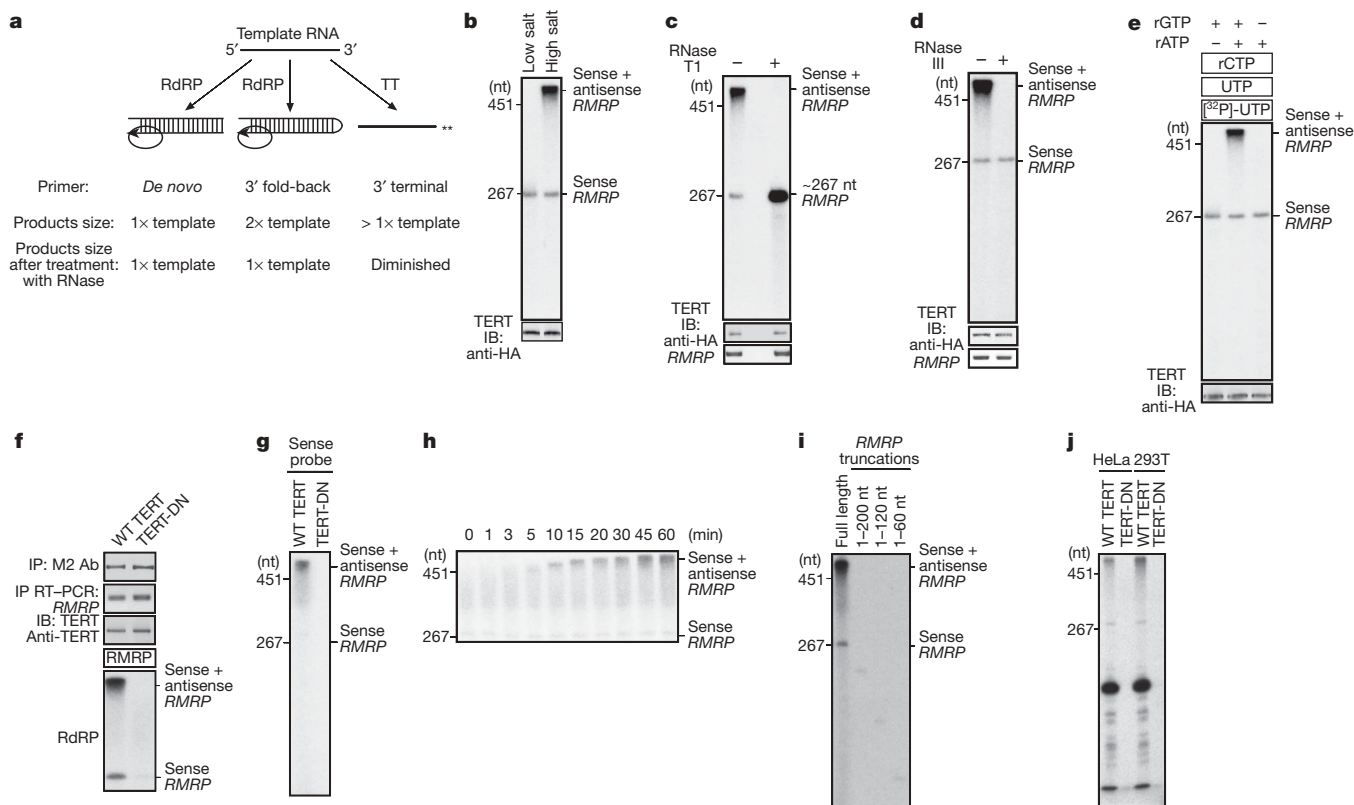


Figure 2 | TERT and *RMRP* have RdRP activity. **a**, Predicted RNA products produced by RdRP or terminal transferase (TT) activity. **b**, RNA products produced by the RdRP activity derived from recombinant TERT and *RMRP*. nt, nucleotides. **c**, **d**, Treatment of RNA products with RNase T1 (**c**) or bacterial RNase III (**d**). **e**, RdRP assay performed in the presence of ribonucleotides (middle) or in the absence of adenine (left lane) or guanine (right lane) ribonucleotides. A and G are present within the first 5 nucleotides of the predicted complementary strand of *RMRP*. **f**, TERT-DN binds *RMRP* but lacks RdRP activity. TERT immune complexes were

isolated from 293T cells expressing Flag-tagged TERT or Flag-tagged TERT-DN. RdRP activity is shown in the bottom panel. WT, wild-type. **g**, Northern blotting to detect complementary sequence of *RMRP*. **h**, Time course of RdRP activity. **i**, RNA products produced by recombinant TERT and truncation mutants of *RMRP* transcribed *in vitro*. Faint signals at 200, 120 and 60 nucleotides are TERT terminal transferase products. **j**, RNA products produced by the RdRP activity derived from recombinant TERT or TERT-DN and total RNA. A limited pool of RNAs serves as template for RdRP activity.

combinations of recombinant mutant TERT proteins and *RMRP*. We failed to detect RdRP reaction products when TERT and *TERC* were co-incubated (Supplementary Fig. 8). Moreover, when we used the TERT-HT1 mutant that does not bind *RMRP* (Supplementary Fig. 4), we failed to observe labelled RNA products (Supplementary Fig. 8) under conditions in which we detected two different RNA products in reactions containing wild-type TERT and *RMRP*. We previously described a catalytically inactive TERT mutant (TERT-DN) that fails to elongate telomeres^{11,27}. We confirmed that the recombinant TERT-DN mutant retained the ability to bind *RMRP* (Fig. 2f), but that the TERT-DN-*RMRP* complex lacked detectable RdRP activity (Fig. 2f). Thus TERT acts as the catalytic subunit for both the telomerase reverse transcriptase and RdRP activities.

TERT-*RMRP* RdRP produces dsRNA

These observations suggest that the TERT-*RMRP* RdRP synthesizes dsRNA in a template-dependent manner. To confirm the synthesis of the *RMRP* complementary strand, we used the sense strand of *RMRP* as a probe in northern blotting. We detected the antisense strand of *RMRP* in reactions containing recombinant wild-type TERT protein and *RMRP*, but not in reactions containing TERT-DN and *RMRP* (Fig. 2g). Furthermore, we detected the sense plus antisense product in the RdRP assay using the antisense strand of *RMRP* as a probe (Supplementary Fig. 9). These observations indicate that the TERT-*RMRP* RdRP produces dsRNAs in a template-dependent manner *in vitro*.

To determine whether the TERT-*RMRP* RdRP uses a back-priming mechanism, we examined the priming process using TERT and *RMRP* as a model system and found that elongation products appeared in a time-dependent manner (Fig. 2h and Supplementary Fig. 10). To assess whether the *RMRP* RNA forms a 3' fold-back configuration, we generated 3' *RMRP* truncation mutants and failed to find any reaction products (Fig. 2i). Thus, unlike what has been described for other cellular RdRPs, the TERT-*RMRP* RdRP has a restricted preference for RNA molecules that can be used as a template. Indeed, when we incubated purified recombinant TERT together with total cellular RNA and [³²P]UTP, we identified a limited number of labelled RNAs (Fig. 2j). Although the secondary structure adopted by *RMRP* to create the 3' fold-back is not known, these observations suggest that *RMRP* can itself serve as a primer for the polymerization process using a 3' fold-back structure.

To ascertain whether this RdRP activity also occurs *in vivo*, we used the sense strand of *RMRP* as a probe and found ~534-nucleotide RNAs that contain antisense *RMRP* in RNA derived from 293T, HeLa and MCF7 cells (Fig. 3a and Supplementary Figs 11 and 12). Moreover, we detected sense products and sense plus antisense products using *RMRP* antisense-strand probe (Fig. 3b). These observations confirmed that the ~534-nucleotide products contain both sense and antisense *RMRP* sequences. To determine whether TERT was necessary for the appearance of antisense *RMRP* in cells, we examined the levels of the complementary *RMRP* strand in cells that do not express TERT and *TERC* (VA-13 cells)²⁸, in cells that transiently express low levels of TERT (BJ cells)^{27,29,30}, and in cells that constitutively express TERT (293T and HeLa cells). We also introduced a control vector or a vector that encodes TERT in VA-13 and BJ cells. We detected the complementary *RMRP* strand using a quantitative RNase protection assay with a sense-strand probe that detects antisense *RMRP* (Fig. 3c and Supplementary Fig. 13), and using northern blotting with both sense and antisense strand-specific *RMRP* probes (Fig. 3d and Supplementary Fig. 11a). The levels of antisense *RMRP* correlated with the expression of TERT (Fig. 3c, d). These observations confirmed that the TERT-*RMRP* RdRP produces double-stranded *RMRP* *in vivo*.

Effects of the TERT-*RMRP* complex on *RMRP* expression

To assess the consequences of overexpressing the TERT-*RMRP* complex on *RMRP* levels, we introduced *RMRP* into cells that lack TERT

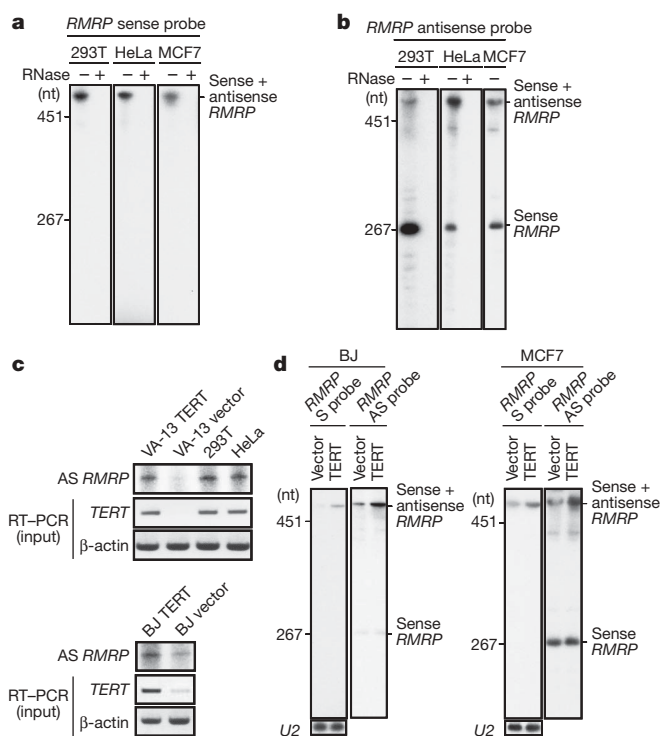


Figure 3 | Identification of dsRNA synthesized by the TERT-*RMRP* RdRP.

a, Northern blotting to detect complementary sequence of *RMRP* in cell lines. '+' indicates samples treated with RNase. **b**, Northern blotting to detect the *RMRP* sense strand. **c**, TERT expression correlates with the levels of antisense (AS) *RMRP* detected by RNase protection assay. Vector denotes cells infected with a control vector. **d**, TERT expression correlates with the levels of the sense (S) plus antisense *RMRP* products detected by northern blotting. The bottom panel shows levels of the small nuclear RNA U2.

expression (VA-13), that transiently express TERT in a cell-cycle-dependent manner (BJ fibroblasts), and that constitutively express TERT (VA-13 and BJ fibroblasts expressing ectopic TERT, and HeLa and MCF7 cells). After expressing *RMRP* in cells lacking TERT (VA-13), we found that *RMRP* levels were increased (Fig. 4a and Supplementary Fig. 14). In contrast, in cells that express TERT, we found that the steady-state levels of *RMRP* were decreased when *RMRP* was overexpressed, regardless of the promoter that was used to express *RMRP* (Fig. 4a and Supplementary Fig. 14). We also found that forced TERT expression in VA-13 or BJ cells suppressed *RMRP* expression (Fig. 4b and Supplementary Fig. 15). Consistent with these findings, suppression of TERT in HeLa cells led to increased *RMRP* expression (Fig. 4c).

Because the 3' end of *RMRP* was essential for TERT-*RMRP* activity (Fig. 2i), we examined the effects of expressing *RMRP* truncation mutants lacking 3' ends and found that only truncation mutants lacking intact 3' ends were readily overexpressed (Fig. 4d). These observations demonstrate that *RMRP* expression levels are dependent on the TERT-*RMRP* RdRP and suggest that *RMRP* levels are controlled by an RdRP-dependent, negative-feedback mechanism.

Identification of siRNAs derived from *RMRP*

In other organisms, RdRPs synthesize dsRNAs that are processed into active short interfering RNAs (siRNAs)³¹. Because manipulating TERT and *RMRP* levels affected *RMRP* expression, we proposed that the TERT-*RMRP* complex produces *RMRP*-specific siRNA to regulate *RMRP* levels. To test this possibility, we used sense and antisense probes corresponding to *RMRP* (nucleotides 21–40) in northern blotting and found double-stranded 22-nucleotide RNAs (Fig. 4e and Supplementary Fig. 11b). Because siRNAs contain 5' monophosphate and 3' hydroxyl groups^{32–34}, we characterized the chemical

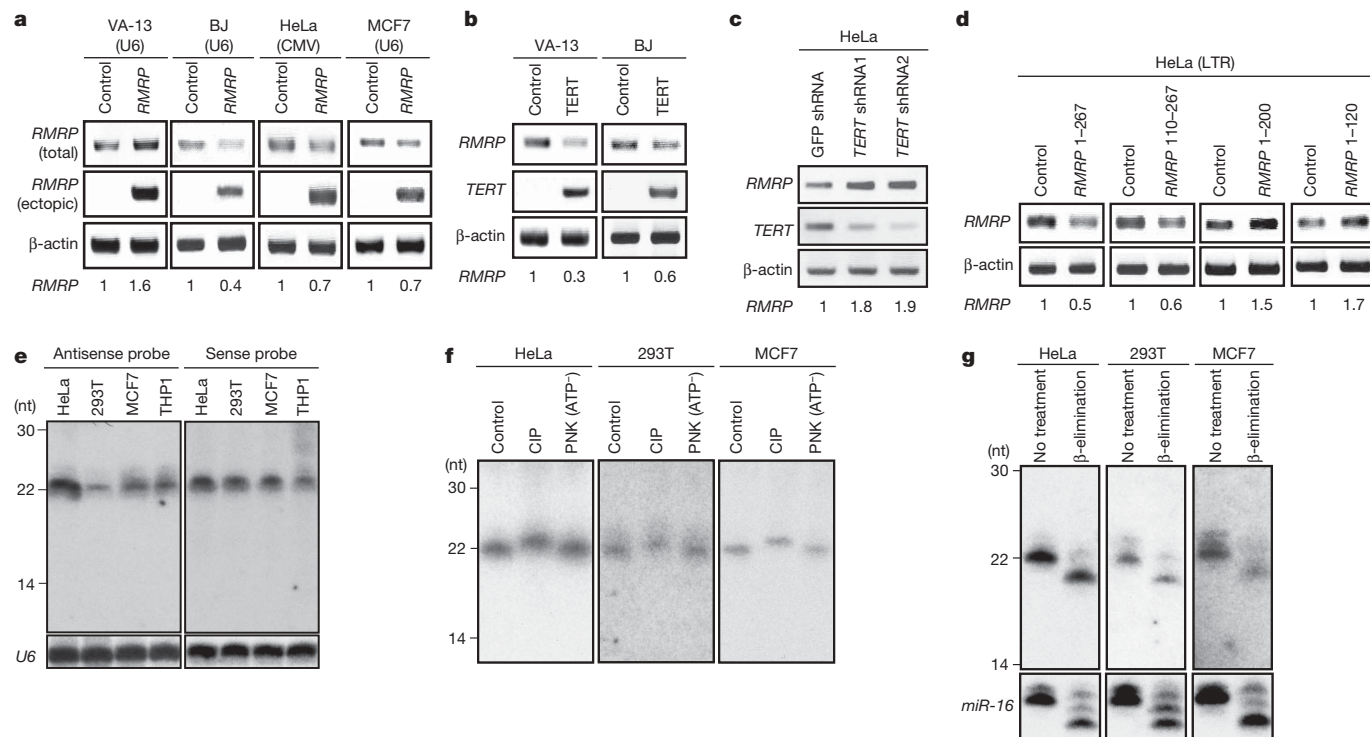


Figure 4 | Effects of dsRNA produced by the TERT–RMRP RdRP. **a**, Semi-quantitative RT–PCR for total *RMRP* and retrovirally delivered *RMRP* (ectopic) in cell lines expressing control or *RMRP* expression vectors. Promoters used to express *RMRP* are indicated. The relative intensity of *RMRP* is noted below each panel. CMV, cytomegalovirus. See Supplementary Fig. 14. **b**, RT–PCR for total *RMRP*. See Supplementary Fig. 15. **c**, Effects of suppressing TERT on *RMRP* levels. A control shRNA (green fluorescent protein (GFP) shRNA) or two different *TERT*-specific shRNAs were stably introduced into HeLa cells. **d**, Effects of *RMRP* mutants on

RMRP levels. LTR, long terminal repeat. RT–PCR was used to detect *RMRP* levels in **c** and **d**. **e**, Detection of small RNA species in human cells. Northern blotting to detect small RNAs (22 nucleotides in length) using antisense (left panel) and sense (right panel) probes derived from nucleotides 21–40 of *RMRP*. **f**, **g**, Analysis of the termini of the small RNA species identified in **e**. Total RNA was incubated with the indicated enzyme (**f**), or oxidation- β -elimination reactions (**g**) were performed. Northern blotting was performed with antisense probe. CIP, calf intestinal phosphatase; PNK, polynucleotide kinase. ATP- indicates samples lacking ATP.

nature of the small RNA ends. We found that calf intestinal phosphatase slowed the migration of these short RNAs, and subsequent incubation with polynucleotide kinase and ATP restored the mobility of the short RNAs, indicating that either the 5' or the 3' end of this small RNA is monophosphorylated (Fig. 4f and data not shown). Moreover, incubation with polynucleotide kinase in the absence of ATP did not alter the migration (Fig. 4f), and oxidation and β -elimination treatment increased the migration of these small RNAs (Fig. 4g), indicating that the 3' ends bear vicinal 2',3' dihydroxyls. Together, these observations confirm that these small RNAs contain 5' monophosphate and 3' hydroxyl groups, and therefore share the size and chemical composition of known siRNAs.

To demonstrate that dsRNAs produced by the TERT–RMRP RdRP are processed into siRNA, we suppressed the expression of *Dicer* with two distinct *Dicer*-specific short hairpin RNAs (shRNAs). Suppression of *Dicer* to levels that partially inhibited the processing of the microRNA *miR-16* (Fig. 5a and Supplementary Fig. 16) led to diminished levels of the siRNAs derived from *RMRP* (Fig. 5a). When we suppressed *Dicer* expression in HeLa, 293T or MCF7 cells, we found that endogenous *RMRP* levels increased up to 3.7-fold (Fig. 5b). Suppressing *Dicer* expression in VA-13 cells that lack TERT did not affect the levels of single-stranded *RMRP* (Fig. 5b), but did increase levels of the elongated sense plus antisense *RMRP* products in cells that constitutively express TERT (Supplementary Fig. 17). Moreover, we found that only the sense strands of these endogenous *RMRP*-specific siRNAs were associated with human AGO2 (also known as EIF2C2; Fig. 5c). These observations indicate that the endogenous *RMRP*-specific siRNAs are processed by the RNA-induced silencing complex, similar to other small RNAs that are processed into siRNA.

To confirm that these small RNAs act as siRNAs, we identified small RNAs from total RNA that hybridized to probes spanning *RMRP*, synthesized siRNA corresponding to the identified sequences, and tested the consequences of introducing this siRNA in HeLa, 293T and MCF7 cells. We found that the synthesized siRNA suppressed endogenous *RMRP* levels (Supplementary Fig. 18). These observations provide evidence that similar to other cellular RdRPs, the TERT–RMRP RdRP synthesizes dsRNAs that act as a precursor for siRNAs.

Discussion

Here we demonstrate that human TERT and *RMRP* form a distinct ribonucleoprotein complex that has the ability to produce dsRNAs (Supplementary Fig. 1). Like RdRPs found in other organisms, the human TERT–RMRP complex produces dsRNAs that act as substrates for the generation of siRNA. However, unlike other cellular RdRPs^{23,26,31,35,36}, the human TERT–RMRP RdRP shows a strong preference for RNA templates that can form 3' fold-back structures. Because other cellular RdRPs have been identified using assays that require primer-independent RdRP activity^{23,26,36}, the substrate specificity of the human TERT–RMRP RdRP may, in part, account for the difficulty in identifying mammalian enzymes that have RdRP activity.

Although the cellular RdRPs described until now do not show a primer requirement, several viral RdRPs use both primer-dependent and primer-independent mechanisms, and fungal and yeast RdRPs are also able to use a back-priming mechanism^{23,26}. Because TERT is a closed right-handed polymerase³⁷ evolutionarily related to both reverse transcriptases and viral RdRPs¹⁸, these observations are consistent with previous observations that indicate that right-handed RdRPs exhibit primer-dependent RdRP polymerase activity³⁸.

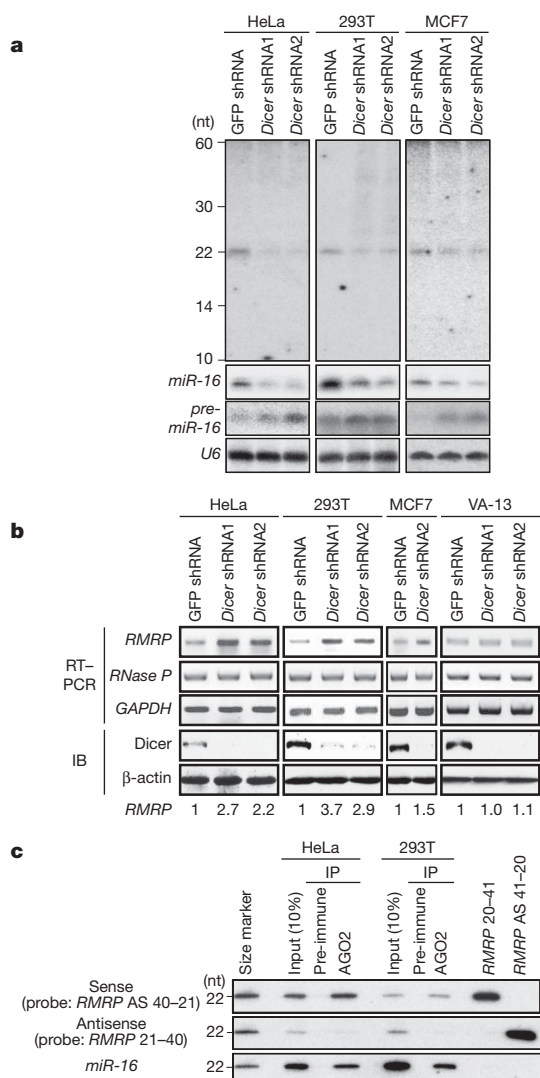


Figure 5 | Production of RMRP-derived endogenous siRNAs depends on Dicer. **a**, Effect of suppressing *Dicer* on RMRP-derived small RNAs. Northern blotting was performed to detect: (1) small RNAs using the antisense strand of RMRP as a probe in the indicated cells expressing control shRNA (GFP shRNA) or *Dicer*-specific shRNAs (*Dicer* shRNA1 and shRNA2); (2) precursor microRNA *pre-miR-16* and mature *miR-16* using a *miR-16*-specific probe; and (3) *U6* RNA. See Supplementary Fig. 16. **b**, RT-PCR for total RMRP from cell lines expressing control shRNA or *Dicer*-specific shRNAs. IB, immunoblot. The relative intensity of RMRP is noted at the bottom of the panel. **c**, RMRP-derived small RNAs are associated with AGO2. Human AGO2 immune complexes were isolated using anti-AGO2-specific antisera or pre-immune sera, and small RNAs were detected by northern blotting. Blotting of oligonucleotides (RMRP 20–41 and RMRP AS 41–20) is also shown.

Using RMRP as a template, the TERT–RMRP RdRP produces dsRNAs that are processed by Dicer into 22-nucleotide dsRNAs that contain 5' monophosphate and 3' hydroxyl groups and are loaded into AGO2, confirming that these short RNAs represent endogenous siRNAs. Recent work has shown that in oocytes and embryonic stem cells, endogenous siRNA can also be formed by the transcription of complementary sense and antisense strands^{39–41}. Thus, in mammals at least two mechanisms lead to the production of dsRNAs that are processed into siRNA. Further work will be necessary to determine whether there are tissue-dependent differences in the use of these two mechanisms and whether other mammalian RdRPs exist.

We found that the TERT–RMRP RdRP regulates RMRP levels by a negative-feedback control mechanism. The identities and functions

of the RNAs other than RMRP that act as templates for the TERT–RMRP RdRP remain to be identified (Fig. 2j). However, because endogenously encoded siRNAs suppress L1 retrotransposition in human cells⁴², these observations suggest that the TERT–RMRP complex may regulate the expression of other genes by generating siRNAs.

Because mutations in RMRP are found in cartilage–hair hypoplasia¹⁵, these findings suggest that perturbation of the TERT–RMRP complex is involved in the pathogenesis of this disorder. The involvement of human TERT in two syndromes characterized by stem-cell failure (cartilage–hair hypoplasia and dyskeratosis congenita)^{7,8,43} suggests that ribonucleoprotein complexes containing TERT has a critical role in stem cell biology. Indeed, overexpression of mouse TERT in mice lacking *Terc* leads to defects in normal hair follicle stem-cell function¹² at least in part by altering gene expression programs related to stem cell function⁴⁴. In mammals, TERT may regulate both telomere biology and gene expression through these two ribonucleoprotein complexes.

METHODS SUMMARY

RNAs that bind TERT were identified from HeLa S3 cells expressing a TAP epitope-tagged TERT. RNAs that bound to TERT after two rounds of purification were analysed using an Experion capillary electrophoresis device (Bio-Rad) to visualize RNA species. For RNA cloning and sequencing, the same samples were separated using a 7 M urea/15% polyacrylamide gel, and RNAs recovered from the gel were cloned using a small RNA cloning Kit (TaKaRa). Purified glutathione S-transferase (GST)–TERT was isolated from *Escherichia coli* and incubated with either *TERC* or RMRP transcribed *in vitro*, to assess the ability of such complexes to exhibit telomerase or RdRP activity. RNAi was used to suppress TERT and to show that the TERT–RMRP complex also produces dsRNA in cells. Northern blotting with sense and antisense probes specific for RMRP (nucleotides 21–40) identified 22-nucleotide, double-stranded RNAs that contained a 5' monophosphate and a 3' hydroxyl group, which were loaded into human AGO2. To determine the function of these RMRP-derived small RNAs, a chemically synthesized siRNA corresponding to these small RNAs (siRNA: 5'-GGCTACACACTGAGGACTC-3'; Dharmacon) was transfected into HeLa, 293T and MCF7 cells.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 May; accepted 10 July 2009.

Published online 23 August 2009.

- Cohen, S. B. *et al.* Protein composition of catalytically active human telomerase from immortal cells. *Science* **315**, 1850–1853 (2007).
- Fu, D. & Collins, K. Purification of human telomerase complexes identifies factors involved in telomerase biogenesis and telomere length regulation. *Mol. Cell* **28**, 773–785 (2007).
- Venteicher, A. S., Meng, Z., Mason, P. J., Veenstra, T. D. & Artandi, S. E. Identification of ATPases pontin and reptin as telomerase components essential for holoenzyme assembly. *Cell* **132**, 945–971 (2008).
- Venteicher, A. S. *et al.* A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. *Science* **323**, 644–648 (2009).
- Weinrich, S. L. *et al.* Reconstitution of human telomerase with the template RNA component hTR and the catalytic protein subunit hTERT. *Nature Genet.* **17**, 498–502 (1997).
- Chan, S. W. & Blackburn, E. H. New ways not to make ends meet: telomerase, DNA damage proteins and heterochromatin. *Oncogene* **21**, 553–563 (2002).
- Liu, J. M. & Ellis, S. R. Ribosomes and marrow failure: coincidental association or molecular paradigm? *Blood* **107**, 4583–4588 (2006).
- Calado, R. T. & Young, N. S. Telomere maintenance and human bone marrow failure. *Blood* **111**, 4446–4455 (2008).
- Gonzalez-Suarez, E. *et al.* Increased epidermal tumors and increased skin wound healing in transgenic mice overexpressing the catalytic subunit of telomerase, mTERT, in basal keratinocytes. *EMBO J.* **20**, 2619–2630 (2001).
- Artandi, S. E. *et al.* Constitutive telomerase expression promotes mammary carcinomas in aging mice. *Proc. Natl Acad. Sci. USA* **99**, 8191–8196 (2002).
- Masutomi, K. *et al.* The telomerase reverse transcriptase regulates chromatin state and DNA damage responses. *Proc. Natl Acad. Sci. USA* **102**, 8222–8227 (2005).
- Sarin, K. Y. *et al.* Conditional telomerase induction causes proliferation of hair follicle stem cells. *Nature* **436**, 1048–1052 (2005).
- Lee, J. *et al.* TERT promotes cellular and organismal survival independently of telomerase activity. *Oncogene* **27**, 3754–3760 (2008).

14. Tollervey, D. & Kiss, T. Function and synthesis of small nucleolar RNAs. *Curr. Opin. Cell Biol.* **9**, 337–342 (1997).
15. Ridanpaa, M. *et al.* Mutations in the RNA component of RNase MRP cause a pleiotropic human disease, cartilage-hair hypoplasia. *Cell* **104**, 195–203 (2001).
16. Moriarty, T. J., Huard, S., Dupuis, S. & Autexier, C. Functional multimerization of human telomerase requires an RNA interaction domain in the N terminus of the catalytic subunit. *Mol. Cell. Biol.* **22**, 1253–1265 (2002).
17. Lue, N. F. *et al.* Telomerase can act as a template- and RNA-independent terminal transferase. *Proc. Natl Acad. Sci. USA* **102**, 9778–9783 (2005).
18. Nakamura, T. M. *et al.* Telomerase catalytic subunit homologs from fission yeast and human. *Science* **277**, 955–959 (1997).
19. Mourrain, P. *et al.* *Arabidopsis* SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell* **101**, 533–542 (2000).
20. Smardon, A. *et al.* EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*. *Curr. Biol.* **10**, 169–178 (2000).
21. Lipardi, C., Wei, Q. & Paterson, B. M. RNAi as random degradative PCR: siRNA primers convert mRNA into dsRNAs that are degraded to generate new siRNAs. *Cell* **107**, 297–307 (2001).
22. Sijen, T. *et al.* On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* **107**, 465–476 (2001).
23. Makeyev, E. V. & Bamford, D. H. Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes. *Mol. Cell* **10**, 1417–1427 (2002).
24. Semler, B. L. & Wimmer, E. *Molecular Biology of Picornaviruses* 255–267 (American Society for Microbiology, 2002).
25. Behrens, S. E., Tomei, L. & De Francesco, R. Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus. *EMBO J.* **15**, 12–22 (1996).
26. Sugiyama, T., Cam, H., Verdel, A., Moazed, D. & Grewal, S. I. RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proc. Natl Acad. Sci. USA* **102**, 152–157 (2005).
27. Masutomi, K. *et al.* Telomerase maintains telomere structure in normal human cells. *Cell* **114**, 241–253 (2003).
28. Ford, L. P. *et al.* Telomerase can inhibit the recombination-based pathway of telomere maintenance in human cells. *J. Biol. Chem.* **276**, 32198–32203 (2001).
29. Pascale, E., Cimino Reale, G. & D'Ambrosio, E. Tumor cells fail to *trans*-induce telomerase in human umbilical vein endothelial cell cultures. *Cancer Res.* **64**, 7702–7705 (2004).
30. Won, J., Chang, S., Oh, S. & Kim, T. K. Small-molecule-based identification of dynamic assembly of E2F-pocket protein-histone deacetylase complex for telomerase regulation in human cells. *Proc. Natl Acad. Sci. USA* **101**, 11328–11333 (2004).
31. Almeida, R. & Allshire, R. C. RNA silencing and genome regulation. *Trends Cell Biol.* **15**, 251–258 (2005).
32. Schwarz, D. S., Hutvagner, G., Haley, B. & Zamore, P. D. Evidence that siRNAs function as guides, not primers, in the *Drosophila* and human RNAi pathways. *Mol. Cell* **10**, 537–548 (2002).
33. Schwarz, D. S., Tomari, Y. & Zamore, P. D. The RNA-induced silencing complex is a Mg²⁺-dependent endonuclease. *Curr. Biol.* **14**, 787–791 (2004).
34. Vagin, V. V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320–324 (2006).
35. Nishikura, K. A short primer on RNAi: RNA-directed RNA polymerase acts as a key catalyst. *Cell* **107**, 415–418 (2001).
36. Aoki, K., Moriguchi, H., Yoshioka, T., Okawa, K. & Tabara, H. *In vitro* analyses of the production and activity of secondary small interfering RNAs in *C. elegans*. *EMBO J.* **26**, 5007–5019 (2007).
37. Gillis, A. J., Schuller, A. P. & Skordalakes, E. Structure of the *Tribolium castaneum* telomerase catalytic subunit TERT. *Nature* **455**, 633–637 (2008).
38. Salgado, P. S. *et al.* The structure of an RNAi polymerase links RNA silencing and transcription. *PLoS Biol.* **4**, e434 (2006).
39. Tam, O. H. *et al.* Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**, 534–538 (2008).
40. Watanabe, T. *et al.* Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**, 539–543 (2008).
41. Babiarz, J. E., Ruby, J. G., Wang, Y., Bartel, D. P. & Blelloch, R. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev.* **22**, 2773–2785 (2008).
42. Yang, N. & Kazanian, H. H. Jr. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nature Struct. Mol. Biol.* **13**, 763–771 (2006).
43. Guggenheim, R., Somech, R., Grunebaum, E., Atkinson, A. & Roifman, C. M. Bone marrow transplantation for cartilage-hair-hypoplasia. *Bone Marrow Transplant* **38**, 751–756 (2006).
44. Choi, J. *et al.* TERT promotes epithelial proliferation through transcriptional control of a Myc- and Wnt-related developmental program. *PLoS Genet.* **4**, e10 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Sugimura and S. Hirohashi at the National Cancer Center for comments. We also thank H. Siomi, H. Tabara and Y. Tomari for discussions. This work was supported in part by Grant-in-Aid for Young Scientists (A) 19689010 (K.M.) and Grant-in-Aid for Young Scientists (B) 19791141 (Y.M.) from the Ministry of Education, Culture, Sports, Science and Technology, by the Third-Term Comprehensive Control Research for Cancer (K.M.) from the Ministry of Health, Labor, and Welfare, by a Takeda Science Foundation grant (K.M.), the Uehara Memorial Foundation (K.M.), a J&J Focused Funding Award (W.C.H.) and R01 AG23145 from the National Institutes of Health (W.C.H.).

Author Contributions Y.M., M.Y., M.F., N.O., R.P. and V.K. performed experiments. T.L. and Y.H. designed and carried out the bioinformatics analyses of TERT-associated RNAs. Y.M., M.Y., W.C.H. and K.M. designed the experiments and discussed the interpretation of the results. W.C.H. and K.M. cowrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to W.C.H. (william_hahn@dfci.harvard.edu) or K.M. (kmasutom@ncc.go.jp).

METHODS

Cell culture and stable expression of TAP-TERT. The human cell lines 293T, MCF7, HeLa, HeLa S3 and VA-13 were maintained in DMEM supplemented with 10% heat-inactivated FBS. BJ fibroblasts were cultured as described⁴⁵. Amphotropic retroviruses were created as described^{45,46} using the vectors pWZL-Blast-N-Flag/HA-TERT (for HeLa-S3-TAP-TERT), pBABE-puro or pBABE-puro-TERT. After infection, cells were selected with blasticidin S ($10 \mu\text{g ml}^{-1}$) for 5 days or with puromycin ($2 \mu\text{g ml}^{-1}$) for 3 days.

Purification of TERT complexes and cloning of RNAs. HeLa S3 cells (2×10^8) expressing or lacking (control) TAP-TERT were lysed in 5 ml of lysis buffer A (20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.5% NP-40, 0.1 mM dithiothreitol (DTT)) and incubated for 30 min on ice. The lysate was then pelleted by centrifugation (16,000g) for 20 min at 4 °C. The supernatant was incubated with anti-Flag (M2) antibody-conjugated agarose overnight at 4 °C. The beads were washed three times with lysis buffer A and eluted with $3 \times$ Flag peptide ($150 \text{ ng } \mu\text{l}^{-1}$). The resulting elution was incubated with Protein A Sepharose beads and an anti-HA antibody (F7; Santa Cruz) for 4 h at 4 °C. The beads were washed three times with lysis buffer A, and RNA was isolated using TRIzol (Invitrogen). RNA samples prepared in this manner were analysed using an Experion capillary electrophoresis device (Bio-Rad) to visualize RNA species. For RNA cloning and the sequencing, the same samples were separated using a 7 M urea/15% polyacrylamide gel, and RNAs recovered from gel were cloned using a small RNA cloning Kit (TaKaRa).

RNA preparation for immunoprecipitation RT-PCR. RNA samples that were prepared from the HeLa S3 cells expressing TAP-TERT as described earlier were also subjected to RT-PCR. For immunoprecipitation of endogenous TERT complexes, cells (1×10^8) were lysed in 600 μl of lysis buffer A, sonicated and pre-cleared with 15 μl of 50% slurry of Protein A Sepharose (Pierce) for 2 h at 4 °C. The pre-cleared total cell lysate was incubated with a rabbit polyclonal anti-TERT antibody (Rockland, 2 μl) for 3 h at 4 °C, followed by incubation with 30 μl of 50% slurry of Protein A Sepharose overnight at 4 °C. After binding, the beads were washed three times for 30 min with lysis buffer A. RNA derived from a single immunoprecipitation was isolated from the Protein A Sepharose using TRIzol (Invitrogen) followed by RT-PCR with primers specific for *TERC*, *RMRP* or *RNase P*. Although other RNAs also co-purified with human TERT (Supplementary Table 1), we failed to confirm the interaction of *Alu* sequences or the 5.8S ribosomal RNA on the Y chromosome with TERT (data not shown).

RT-PCR and quantitative RT-PCR. Either total cellular RNA or RNA from immunoprecipitation was isolated using TRIzol (Invitrogen) and subjected to RT-PCR. The following primers were used: *TERC* (43F, 5'-TCTAAGCC TAAGTGAAGGCGT-3' and 163R, 5'-TGCTCTAGAATGAACGGTGA AGG-3'), *RMRP* (F5, 5'-TGCTGAAGGCCGTATCCT-3' and R257, 5'-TGAGAATGAGCCCCGTGT-3'), *RNase P* (F50, 5'-GTCACTCCACTCC CATGTCC-3' and R318, 5'-AATTGGGTTATGAGGTCCC-3'), and the human β -actin gene (also known as *ACTB*) (5'-CAAGAGATGGCCACGGCTGCT-3' and 5'-TCCTTCTGCATCCTGTCCGCA-3'). The reverse transcription reaction was performed for 60 min at 42 °C using the recovered RNA, and PCR was immediately performed (22 cycles for 293T cells, and 26 cycles for HeLa cells: 94 °C, 30 s; 60 °C, 30 s; 72 °C, 30 s).

Quantitative RT-PCR (qRT-PCR) was performed with a LightCycler 480 II (Roche) according to the manufacturer's protocols. The expression levels of *RMRP* were detected using the following primers and probe; forward primer, 5'-GAGAGTGCACGTGCATACG-3', reverse primer, 5'-CTCAGCGGGATA-CGCTTCTT-3', VIC-labelled TaqMan MGB probe, 5'-ACGTAGACATT-CCCC-3'. β -actin was used as a reference.

Total *RMRP* was detected using primers (F5, 5'-TGCTGAAGGCC TGTATCCT-3' and R257, 5'-TGAGAATGAGCCCCGTGT-3') that amplify both endogenous and ectopically introduced *RMRP*. In Fig. 4a, for VA-13, BJ and MCF7 cells, reverse transcription was performed using random hexamers (GE Healthcare) and ectopically expressed *RMRP* was detected with vector-specific primers (F5, 5'-TGCTGAAGGCCGTGTATCCT-3' and LKO.1-RT, 5'-ACTGCCATTTGTCTCGAGGT-3'). For HeLa cells, reverse transcription was performed with pQC3' (5'-AAGCGGCTTCGGCCAGTAACGTTA-3') and PCR was performed with the primers F5 (5'-TGCTGAAGGCCGTGTATCCT-3') and R257 (5'-TGAGAATGAGCCCCGTGT-3'). Northern blotting and qRT-PCR experiments (Supplementary Fig. 14) confirmed that the differences in *RMRP* levels that were observed using the RT-PCR conditions used in Fig. 4a accurately reflect *RMRP* levels. Signal intensity was measured with ImageJ software.

Telomerase activity reconstituted *in vitro* and TRAP assay. *In vitro* reconstitution of telomerase activity (telomere-specific reverse transcriptase activity) was performed as described previously⁵. In brief, recombinant TERT was expressed in the TnT T7-Coupled Reticulocyte Lysate System (Promega) following the

manufacturer's instructions. Purified *TERC* or *RMRP* was included in the *in vitro* transcription/translation reactions. The telomeric repeat amplification protocol (TRAP)⁴⁵⁻⁴⁷ was used to detect telomere-specific reverse transcriptase activity.

Affinity purification of recombinant GST-TERT fusion proteins. GST-TERT-HA, GST-TERT-HT1 and GST-TERT-DN in the pGENKZ expression vector⁴⁸ were provided by S. Murakami. Bacteria (BL21-Gold) containing these vectors were plated at 30 °C overnight and then a single colony was picked to inoculate liquid cultures, which were incubated at 37 °C overnight. Thereafter 1 ml of this culture was re-inoculated into 100 ml of Luria-Bertani medium, incubated at 37 °C for 4 h without isopropyl- β -D-thiogalactoside (IPTG) induction, collected by centrifugation, suspended in a lysis buffer (20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.5% NP-40, 0.1 mM DTT, 10 mM phenylmethyl sulphonyl fluoride (PMSF), proteinase inhibitor (Nacalai Tesque)) and sonicated twice for 10 s at 4 °C. After centrifugation of the sonicated lysates, the supernatants were passed through DEAE-Sepharose, and the GST-fusion proteins were recovered using glutathione-Sepharose 4B beads. The resin was washed with lysis buffer A at least three times, and the GST-fusion proteins were then eluted with glutathione at 4 °C for 1 h (20 mM glutathione (reduced form)) in elution buffer (50 mM Tris-HCl, pH 8.8, 150 mM NaCl, 0.5% NP-40, 0.1 mM DTT, 10 mM PMSF, proteinase inhibitor (Nacalai Tesque)). Supplementary Fig. 6 shows that wild type and TERT-DN were produced at similar levels using this method and the effects of incubation time and IPTG on yield. The average yield for this method is 500 ng ($5 \text{ ng } \mu\text{l}^{-1}$) of active form of TERT from 100 ml culture.

RdRP assay. The affinity purified recombinant GST-TERT fusion protein (10 ng) was incubated with 1 μg of full length *RMRP* RNA or truncated *RMRP* products (*RMRP* 1–200, *RMRP* 1–120 and *RMRP* 1–60 for Fig. 2i) transcribed *in vitro* (SP6) in 200 mM KCl, 50 mM Tris-HCl, pH 8.3, 10 mM DTT, 30 mM MgCl₂, 50 μM rATP, 50 μM rGTP, 50 μM rCTP and 2 μCi of [α -³²P]UTP at 32 °C for 2 h. To perform the experiments under low salt conditions, 20 μl of 0.2 \times SSC was then added to adjust final salt concentration to 15 mM NaCl and 1.5 mM sodium citrate, whereas 20 μl of 4 \times SSC was added to adjust final salt concentration to 300 mM NaCl and 30 mM sodium citrate to achieve high salt conditions. These mixtures were incubated at 37 °C for a further 1 h. Resulting products were treated with proteinase K to stop the reaction and purified with phenol-chloroform. To ensure that RNA products were completely denatured, we performed both conventional formamide treatment (with 95% formamide/20 mM EDTA gel-loading buffer at 95 °C for 5 min) and a further treatment with 1 M de-ionized glyoxal at 65 °C for 15 min.

To analyse double-stranded RNA produced by the TERT-*RMRP* complex, we performed this RdRP assay and treated the products with bacterial RNase III (*E. coli*, Ambion; 50 mM NaCl, 10 mM Tris-HCl, pH 7.9, 1 mM DTT, 10 mM MgCl₂) or RNase T1 (Roche; 50 mM Tris-HCl, pH 8.3, 300 mM NaCl and 30 mM sodium citrate).

Northern blotting. Total RNA and small RNAs (<200 nucleotides in length) were isolated using a mirVana miRNA Isolation Kit (Ambion) according to the manufacturer's protocol. Total RNA or small RNA (10 μg) was separated on denaturing polyacrylamide gels, then blotted onto Hybond-N+ membranes (GE Healthcare) using a Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad). Hybridization was performed in Church buffer (0.5 M NaHPO₄, pH 7.2, 1 mM EDTA and 7% SDS) containing 10^6 c.p.m. ml⁻¹ of each ³²P-labelled probe for 14 h. The membranes were washed in 2 \times SSC, and the signals were detected by autoradiography.

Identification of short RNA species derived from *RMRP*. Using ten consecutive probes corresponding to the *RMRP* sequence, we found that the small RNAs derived from *RMRP* shown in Figs 4e–g and 5a were detected by probes containing the complementary sequences to nucleotides 21–40 of *RMRP*. To determine the function of these *RMRP*-derived small RNAs, we purchased a chemically synthesized siRNA targeting this 20-nucleotide portion of the *RMRP* sequence (siRNA: 5'-GGCTACACACTGAGGACTC-3'; Dharmacon) and transfected this siRNA into HeLa, 293T and MCF7 cells plated on six-well dishes using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol.

RNase protection assay. *RMRP* RNA was transcribed with SP6 RNA polymerase in the presence of [α -³²P]UTP using RiboMAX Large Scale RNA Production System (Promega). Total cellular RNA (30 μg) was hybridized overnight at 60 °C with equal amounts of ³²P-labelled *RMRP* sense probe. Hybrids were digested with RNase A and RNase T1. The protected fragments were separated by PAGE under denaturing conditions and visualized by autoradiography.

Analysis of the chemical structure of the ends of small RNAs. To determine the phosphorylation status of the termini of small RNAs, 30 μg of small RNA (<200 nucleotides in length) was treated with calf intestinal alkaline phosphatase (CIP; TaKaRa) for 2 h at 37 °C. CIP was inactivated by phenol-chloroform extraction. Part of the CIP-treated RNA was then treated with T4 polynucleotide

kinase (TaKaRa) supplemented with 1 mM ATP for 2 h at 37 °C, and phenol–chloroform extraction was performed. Small RNA (15 µg) was treated with T4 polynucleotide kinase without ATP for 2 h at 37 °C. The reaction was inactivated by phenol–chloroform extraction. After overnight sodium acetate–ethanol precipitation at –20 °C, the treated RNAs were resolved by 20% denaturing polyacrylamide/urea gel electrophoresis and then analysed by northern blotting^{42,43}.

To further analyse the 3' end of these small RNAs, we performed oxidation and β -elimination reactions. Specifically, the NaIO₄ reaction was performed by adding 20 µg of small RNA in water to 5× borate buffer (148 mM borax and 148 mM boric acid, pH 8.6) and freshly dissolved 200 mM NaIO₄ to create a final concentration of 1× borate buffer and 25 mM NaIO₄. The mixtures were incubated for 10 min at 20 °C. Glycerol was added to quench remaining NaIO₄, and the samples were incubated for a further 10 min at 20 °C. For β -elimination, small RNAs were dried by centrifugation and evaporation and dissolved in 50 µl of 1× borax buffer (30 mM borax, 30 mM boric acid and 50 mM NaOH, pH 9.5) and incubated at 45 °C for 90 min. Nucleic acids were recovered by sodium acetate–ethanol precipitation at –20 °C overnight, and the treated RNAs were resolved by 20% denaturing 7 M urea PAGE and analysed by northern blotting⁴³.

Stable expression of shRNA. We used the pLKO.1-puro vector and the sequences described below to create shRNA vectors specific for *TERT*, *Dicer* and GFP. These vectors were used to make amphotropic retroviruses and polyclonal cell populations were purified with selection with puromycin (2 µg ml^{–1}). The sequences used for the indicated short hairpin RNAs are shown below where the capitalized letters represent the targeting sequences: *TERT* shRNA1, 5'-GGAAGACAGTGGTGAACCTCCctcgagGGAAGTTCACCACTGTCTTCCttttt-3' and 5'-aattcaaaaaGGAAGACAGTGGTGAACCTCCctcgagGGAAGTTCACCACTGTCTTCC-3'; *TERT* shRNA2, 5'-GGAACACCAAGAAGTTCATCTctcgagAGATGAACTTCTTGGTGTTCctttt-3' and 5'-aattcaaaaaGGA

ACACCAAGAAGTTCATCTctcgagAGATGAACTTCTTGGTGTTC-3'. *Dicer* sequences: *Dicer* shRNA1, 5'-GCTCGAAATCTTACGCAAATActcgagTATTTCGTAAGATTTTCGAGCttttt-3' and 5'-aattcaaaaaGCTCGAAATCTTACGCAATAActcgagTATTTCGTAAGATTTTCGAGC-3'; *Dicer* shRNA2, 5'-CCACA CATCTTCAAGACTTAActcgagTTAAGTCTTGAAGATGTGTGGttttt-3' and 5'-aattcaaaaaCCACACATCTTCAAGACTTAActcgagTTAAGTCTTGAAGATGTGTGG-3'.

Immunoprecipitation of human AGO2 complexes. HeLa or 293T cells were lysed in lysis buffer A and immunoprecipitation was performed using pre-immune sera or anti-AGO2 antibodies⁴⁹ (provided by H. Siomi and M. C. Siomi). RNA was isolated using TRIzol from the protein A beads and resolved by electrophoresis on 7 M urea 20% PAGE. Small RNAs were detected by northern blotting with an antisense probe, a sense probe derived from nucleotides 21–40 of *RMRP*, or a *miR-16*-specific probe (5'-CGCCAATATTTACGTGCTGCTA-3').

45. Hahn, W. C. *et al.* Creation of human tumour cells with defined genetic elements. *Nature* **400**, 464–468 (1999).
46. Hahn, W. C. *et al.* Inhibition of telomerase limits the growth of human cancer cells. *Nature Med.* **5**, 1164–1170 (1999).
47. Kim, N. W. *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011–2015 (1994).
48. Yamashita, T. *et al.* RNA-dependent RNA polymerase activity of the soluble recombinant hepatitis C virus NS5B protein truncated at the C-terminal region. *J. Biol. Chem.* **273**, 15479–15486 (1998).
49. Azuma-Mukai, A. *et al.* Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. *Proc. Natl Acad. Sci. USA* **105**, 7964–7969 (2008).

LETTERS

The global distribution of pure anorthosite on the Moon

Makiko Ohtake¹, Tsuneo Matsunaga², Junichi Haruyama¹, Yasuhiro Yokota¹, Tomokatsu Morota¹, Chikatoshi Honda³, Yoshiko Ogawa³, Masaya Torii⁴, Hideaki Miyamoto⁵, Tomoko Arai⁶, Naru Hirata³, Akira Iwasaki⁷, Ryosuke Nakamura⁸, Takahiro Hiroi⁹, Takamitsu Sugihara¹⁰, Hiroshi Takeda¹¹, Hisashi Otake¹², Carle M. Pieters⁹, Kazuto Saiki¹³, Kohei Kitazato³, Masanao Abe¹, Noriaki Asada³, Hirohide Demura³, Yasushi Yamaguchi¹⁴, Sho Sasaki¹⁵, Shinsuke Kodama⁸, Junya Terazono³, Motomaro Shirao¹⁶, Atsushi Yamaji¹⁷, Shigeyuki Minami¹⁸, Hiroaki Akiyama¹⁹ & Jean-Luc Josset²⁰

It has been thought that the lunar highland crust was formed by the crystallization and floatation of plagioclase from a global magma ocean^{1,2}, although the actual generation mechanisms are still debated^{2,3}. The composition of the lunar highland crust is therefore important for understanding the formation of such a magma ocean and the subsequent evolution of the Moon. The Multiband Imager⁴ on the Selenological and Engineering Explorer (SELENE)⁵ has a high spatial resolution of optimized spectral coverage, which should allow a clear view of the composition of the lunar crust. Here we report the global distribution of rocks of high plagioclase abundance (approaching 100 vol.%), using an unambiguous plagioclase absorption band recorded by the SELENE Multiband Imager. If the upper crust indeed consists of nearly 100 vol.% plagioclase, this is significantly higher than previous estimates of 82–92 vol.% (refs 2, 6, 7), providing a valuable constraint on models of lunar magma ocean evolution.

The magma ocean hypothesis was proposed on the basis of numerous analyses of lunar samples of ferroan anorthosite² (plagioclase-rich rock with minor amounts of mafic silicates that have a relatively high Fe/Mg ratio) collected from a small portion of the nearside highland regions, although the actual generation mechanisms are still being debated^{2,3}. Therefore, the composition of the lunar highland crust is critical to the investigation of a magma ocean and the subsequent evolution of the Moon. The lateral and vertical rock types of the global crust and their mineral compositions have been investigated in previous studies, which have produced important information using lunar samples², lunar meteorites^{8,9}, remote sensing X-ray¹⁰, γ -ray¹¹ and reflectance spectra^{6,12–14}. New data acquired by the Multiband Imager demonstrated higher spatial resolution of optimized spectral coverage that enables us to acquire a clearer view of the composition of the lunar crust.

Remote-sensing data acquired through visible to near-infrared reflectance spectroscopy with high spatial resolution is one of the best approaches for investigating rock types and their mineral compositions

within the global lunar crust and for examining anorthosite spectra in particular. Anorthosite rocks, which contain plagioclase with trace amounts of iron (>0.1 wt% FeO), exhibit a broad absorption band centred near 1,250 nm owing to the electronic transitions of the minor amounts of Fe²⁺ (ref. 15). The Multiband Imager was designed to detect this anorthosite absorption band.

The Multiband Imager has both visible and near-infrared coverages⁴ with spectral bands at 415, 750, 900, 950 and 1,000 nm (visible) and 1,000, 1,050, 1,250 and 1,550 nm (near-infrared). The instrumental spatial resolution is 20 m (visible) or 62 m (near-infrared) per pixel at the nominal altitude (100 km), which is high compared to the Lunar Prospector gamma-ray and neutron spectrometers (45 km per pixel; ref. 11), and the Clementine UVVIS camera (200 m per pixel on average)^{6,12,13}. Multiband Imager data were calibrated⁴ by adjusting the data from the calibration standard target (the Apollo 16 site) to the laboratory reflectance spectrum of an Apollo 16 soil sample¹⁶. The photometric function proposed by ref. 17 and digital terrain models (DTMs) generated by the Multiband Imager are used to convert the obtained data into the reflectance spectra in the standard viewing geometry. Details of the data analysis procedures are described later in the Methods Summary.

Regional soils suffer heavily from vertical and lateral mixing owing to vigorous cratering processes, so the best locations to examine crustal materials that may not have been subjected to extensive mixing are crater central peaks, walls, ejecta and basin rings^{6,13}. Therefore, we have focused on fresh and nearly regolith-free (mixing-free) locations in craters and basins that have been selected for younger ages and high reflectances. The selected locations are globally distributed over the Moon (Fig. 1 and Supplementary Table 1). These include Jackson (22° N, 197° E), South Ray (−9° N, 15° E), Tycho (−43° N, 349° E), Tsiolkovsky (−20° N, 129° E), and the northern (−10.5° N, 264° E) and eastern portion (−21° N, 274° E) of the Inner Rook Ring of the Orientale basin¹⁸.

¹Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, 3-1-1 Yoshino-dai, Sagami-hara, Kanagawa 229-8510, Japan. ²Center for Global Environmental Research, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki 305-8506, Japan. ³The University of Aizu, Ikki-machi, Aizuwakamatsu, Fukushima 965-8580, Japan. ⁴Fujitsu Ltd., Nakase, Mihama-ku, Chiba, Chiba 261-8588, Japan. ⁵University Museum, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. ⁶Planetary Exploration Research Center, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan. ⁷Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan. ⁸Information Technology Research Institute, National Institute of Advanced Industrial Science and Technology, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan. ⁹Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA. ¹⁰Center for Deep Earth Exploration, Japan Agency for Marine Science and Technology, 3173-25 Showa-machi, Kanazawa-ku, Yokohama, Kanagawa 236-0001, Japan. ¹¹Frontier Plasma Research Center, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan. ¹²Lunar and Planetary Exploration Program Group, Japan Aerospace Exploration Agency, 2-1-1 Sengen, Tsukuba, Ibaraki 305-8505, Japan. ¹³Department of Earth and Space Science, Osaka University, 1-1 Machikaneyama, Toyonaka, Osaka 560-0043, Japan. ¹⁴Department of Earth and Environmental Sciences, Graduate School of Environmental Studies, Nagoya University, D2-1 (510) Furo-cho, Chikusa-ku, Nagoya 464-8601 Japan. ¹⁵RISE Project Office, National Astronomical Observatory of Japan, 2-12 Hoshigaoka, Mizusawa, Oshu 023-0861, Japan. ¹⁶1-3-11, Nishi-Asakusa, Taito-ku, Tokyo 111-0035, Japan. ¹⁷Division of Earth and Planetary Sciences, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan. ¹⁸Department of Electrical Engineering, Osaka City University, Sugimoto 3, Sumiyoshi, Osaka 558-8585, Japan. ¹⁹Intercollegiate cooperation strategy, Wakayama University, 930, Sakaedani, Wakayama 640-8510, Japan. ²⁰Space Exploration Institute CP 774, CH-2002 Neuchâtel, Switzerland.

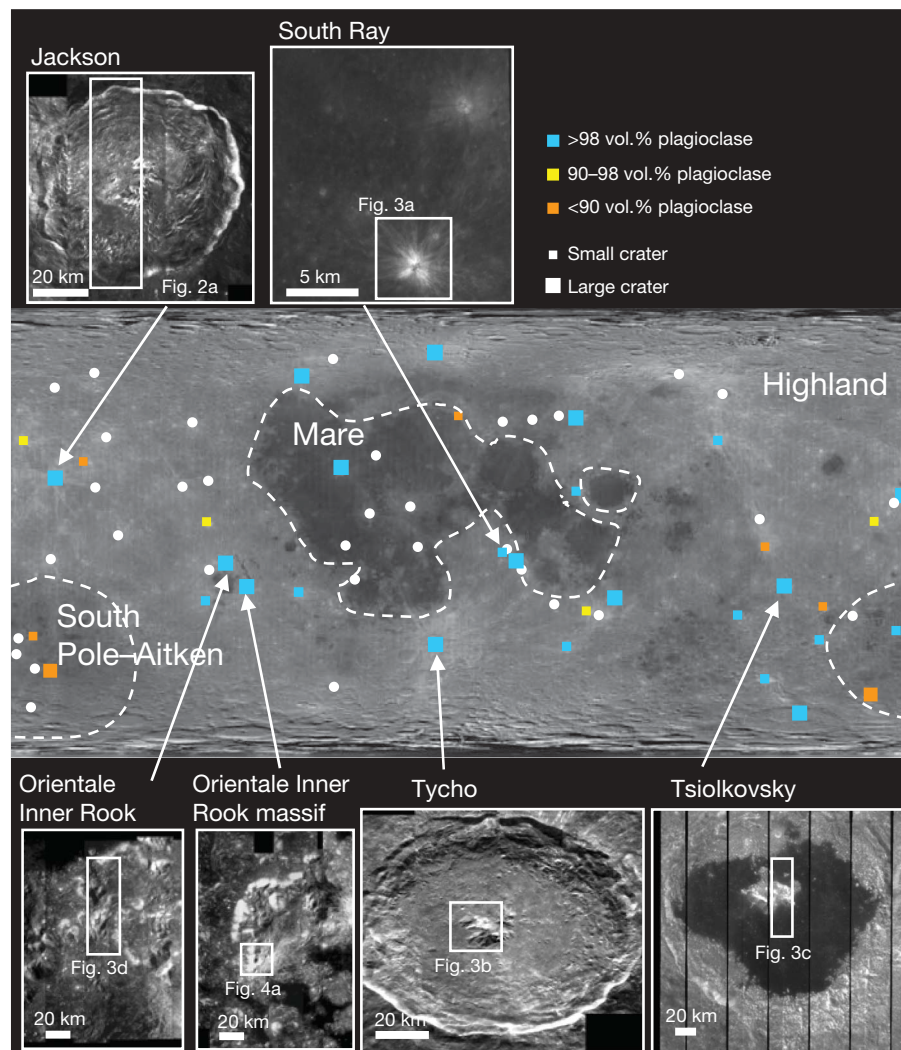


Figure 1 | Locations of the 69 areas of investigation plotted on the USGS Clementine 750-nm basemap. White squares in the inset Multiband Imager mosaics of these areas indicate the exact locations of the images in Figs 2, 3 and 4. Plagioclase modal abundances of the 32 freshest and nearly mixing-free (nearly regolith-free) locations derived from the model analyses are indicated by orange (<90 vol.%), yellow (~90 to ~98 vol.%) and blue (>98 vol.%) squares. The freshest locations are identified using the optical maturity index (≥ 0.6) (ref. 29). Smaller craters (≤ 30 km in diameter) are indicated by small squares, and larger craters are indicated by large squares. Investigated locations that do not have freshly exposed outcrops (optical maturity index <0.6) are plotted as white dots regardless of the crater diameter. PAN rocks sometimes occur in craters with a diameter of up to 30 km but always occur in craters with a diameter exceeding 30 km within highland. South Pole-Aitken is the largest basin on the lunar farside about 2,500 km in diameter.

Jackson (71 km in diameter) is a Copernican-age crater located in the farside highlands. The central peak exhibits a relatively complex structure consisting of brighter slopes and a darker summit, not simple crustal rock (Fig. 2a, b). The angles of the brighter slopes are 33° to 36° , which is equal to or slightly greater than the angle of repose of typical lunar regolith. These brighter slopes are apparently outcrops that have been newly exposed by landslides that occurred much later than the Jackson-generated impact.

In Fig. 2c and d, red, green and blue are assigned to the relative strengths of pyroxene, olivine and plagioclase absorption bands, respectively. The exposed bluish unit is large because the central peak diameter is 7 km and its height is 2.5 km (Fig. 2c, d), suggesting that Fe-bearing crystalline plagioclase is the dominant mineral component based on the prominent 1,250 nm absorption band (J5 and J8 in Fig. 2e, f). Intimate mixing model¹⁹ analyses with plagioclase (FeO content 0.25 wt%), orthopyroxene ($\text{Ca}_3\text{Fe}_{40}\text{Mg}_{57}$), and clinopyroxene ($\text{Ca}_{31}\text{Fe}_{22}\text{Mg}_{47}$) end members (Fig. 2h, i) indicate that this unit is extremely feldspathic (~98 vol.% plagioclase).

Previous researchers proposed a global anorthosite layer deep within the lunar crust. They used restricted data, but were able to derive evidence for the presence of anorthosite in some locations on the Moon using ground-based spectroscopy¹³. However, it was not possible to measure the abundance of plagioclase accurately until the discovery of the 1,250-nm absorption band. Our results enabled us to detect unambiguously an absorption feature generated by trace amounts of iron that is unique to mineral plagioclase and to demonstrate that anorthosite composed of nearly 100% anorthite is found in large exposures. We define a rock that has a high abundance of plagioclase as the purest anorthosite (PAN), to distinguish it from

“pure anorthosite”, which was defined as a rock with over 95 vol.% plagioclase in a previous work¹³. The word ‘pure’ used in the title to refer to anorthosite is meant as a general adjective.

The variation of modal abundance within the brighter unit (blue in Fig. 2c, d) is very small (all areas contain ~98 vol.% plagioclase) in spite of its 7-km diameter. The purity of anorthosite in this entire area is remarkable, considering the generation mechanism for such a massive rock of this purity. PAN rocks also are found at other locations on both sides of the Moon, such as South Ray, Tycho, Tsiolkovsky and Orientale (Fig. 3). These areas exhibit a similar bluish colour in colour-composite images (Fig. 3), which correspond to the prominent 1,250-nm absorption band (Fig. 3e, f). The angles of these bluish inclined planes in Tycho, Tsiolkovsky and Orientale are also as high as in Jackson, suggesting that newly exposed outcrops can be used to spot likely locations of PAN rocks. Intimate mixing models indicate that the modal abundance of plagioclase in these areas is ~98 vol.%, in conformity with the presence of PAN rocks. The occurrence and composition of the relatively mafic-rich units that we identified in Tycho are consistent with the findings of previous studies²⁰, although the PAN rocks are found at the base of the central peak as small (roughly $1\text{ km} \times 2\text{ km}$) outcrops. A relatively large modelled grain size of 400 μm was derived at Tsiolkovsky.

Our finding of a very clear Fe-bearing plagioclase absorption band is in contrast with previous data^{13,21,22}, which lack clear plagioclase absorption (this absence has been an unsolved issue in lunar science¹⁴). The lack of plagioclase absorption in previous spectra has been attributed to shock effects¹³ or space weathering effects²³ known to occur on the Moon. The depth of the Fe-bearing plagioclase absorption band in PAN outcrops at the eastern portion of the

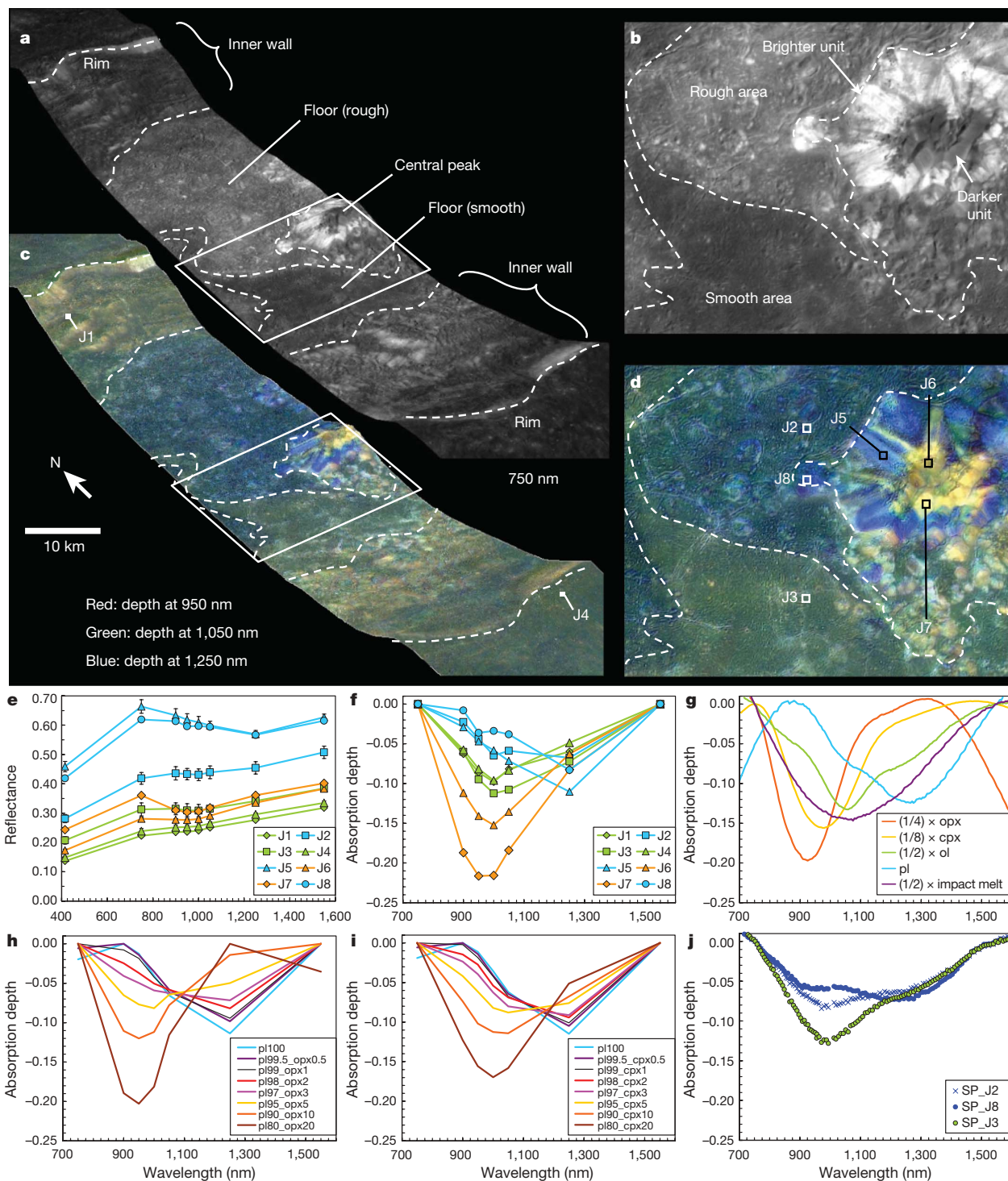


Figure 2 | Results of Jackson analyses. **a**, Bird's-eye-view of Jackson (Multiband Imager 750-nm-band images superimposed on the DTM). The topography is not exaggerated. **b**, Close-up image around the central peak. **c**, Colour-composite bird's-eye-view of the same area as in **a**. Here, red, green and blue are assigned to continuum-removed absorption depths of 950, 1,050 and 1,250 nm, respectively. **d**, Close-up of the colour-composite image. In all images, the spatial resolution is adjusted to 20 m \times 20 m per pixel. **e**, Absolute reflectance spectra at J1 to J8. All the reflectance spectra are given as an average of a 120 m \times 120 m area to remove spatial variation. A standard deviation within the averaged area is presented as an error bar at each data point. **f**, Absorption depths at J1 to J8 versus wavelengths after the continuum removals. **g**, Absorption depths of the lunar impact melt and the minerals separated from the Apollo samples of orthopyroxene (opx), clinopyroxene (cpx), olivine (ol) and plagioclase (pl). Absorption depths of

modelled reflectance spectra in different modal abundances of plagioclase and orthopyroxene (**h**), and plagioclase and clinopyroxene (**i**) mixtures (for example, 'pl97_opx3' means a mixture made up of 97% plagioclase and 3% orthopyroxene). A grain size of 200 μ m was used to generate the observed absorption depth. **j**, Absorption depths at J2, J3 and J8 derived by the Spectral Profiler on SELENE³⁰. The Spectral Profiler (SP) is a line-profiling instrument that provides continuous spectral data. The very high plagioclase abundance (\sim 98 vol.% plagioclase) of the brighter unit (J8) is further confirmed by the Spectral Profiler's independent and continuous spectra. J6 and J7 in **d** are estimated to contain high-Ca pyroxene levels of \sim 10 vol.% and an even higher \geq 20 vol.%, respectively. The origin of the darker unit is unclear; it is probably a mega-regolith layer overlying the crustal material, or impact melt generated by the Jackson impact event that was trapped as the peak was uplifted.

Oriente Inner Rook (the same outcrop studied previously¹³) decreased with decreasing spatial resolution (Fig. 4) owing to the spatial mixing of the reflectance spectra with the spectra of surrounding materials. Moreover, PAN outcrops in some locations are only a few hundred metres in size. Therefore, the Multiband Imager's higher spatial resolution is probably one of the reasons that we were able to detect plagioclase absorption.

Intensive analyses of global Multiband Imager data revealed that PAN rocks are ubiquitous in the central peaks, walls, ejecta, and rings in the global highland (Fig. 1 and Supplementary Table 1). This indicates that the plagioclase absorption band observed in the PAN rocks was not created (or enhanced¹⁵) by a mechanism related to a crater central peak, such as extensive shock deformation and annealing. Anorthosite was

previously suggested to be less abundant or absent in the Procellarum KREEP (potassium, rare-earth elements and phosphorus) Terrane¹². Our results indicate that PAN rocks (defined by modal abundance) are present at one location (Aristarchus) within this terrane, although, judging from the higher Th abundance in the Procellarum KREEP Terrane¹², the generation mechanism of the PAN rocks within the terrane probably differs from that in the highland region.

A strong correlation exists between the distribution of the PAN rocks and the crater size (Fig. 1). PAN rocks occur in all craters under investigation that have diameters exceeding 30 km, whereas less feldspathic rocks occur in some craters with diameters smaller than 30 km. The excavation depth of a 30-km-diameter crater is 3 km. The increased occurrence of PAN rocks at depths exceeding 3 km is possibly due to the inability of small (<30 km in diameter) craters to penetrate surface mega-regolith layers to reach the actual underlying crustal material.

The depth provenances suggested by the diameters of the associated impact craters indicate that PAN rocks are ubiquitously present within the depth range from 3 km to at least 30 km, although some small regional differences may be present (30 km is the original depth of Tsiolkovsky's central peak²⁴). Considering that large, fresh crater central peaks containing a large amount of mafic-rich component are extremely rare in comparison with crater central peaks containing PAN rocks in our study, and the currently estimated average thickness of 27 km (ref. 7) for the upper crust, we believe that a global layer of PAN rock may exist within the upper crust, ranging from 3 km to 30 km in depth. The PAN rocks may exist as large patches within the upper crust. The upper limit for the possible PAN rock layer could be shallower than 3 km because our observation gives only the maximum depth for the upper limit.

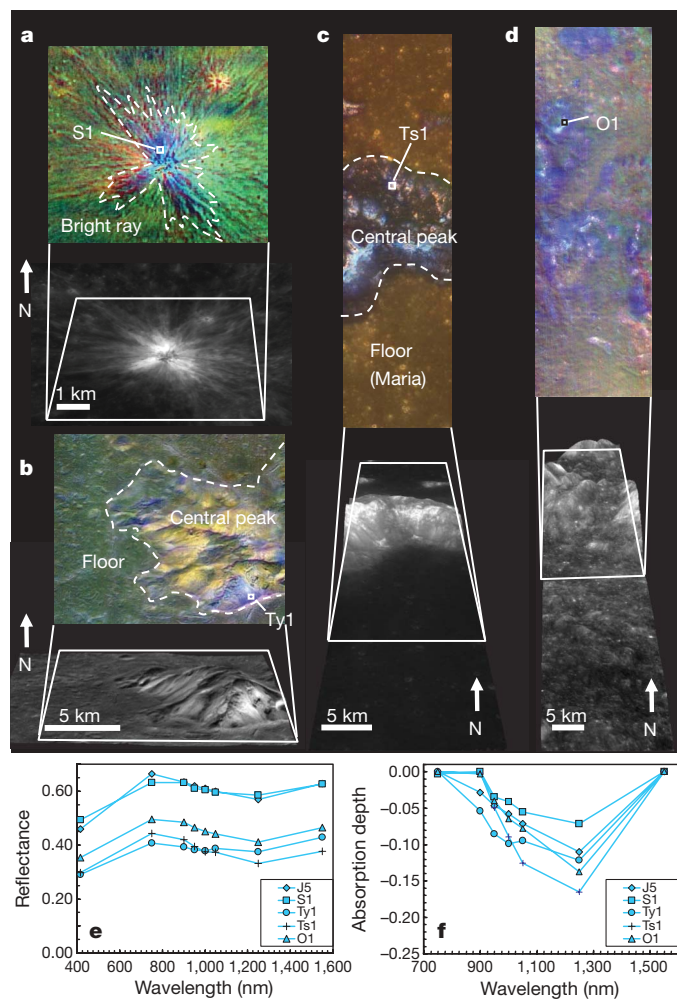


Figure 3 | Results of spatial and spectral analyses of South Ray, Tycho, Tsiolkovsky and Oriente. **a**, Colour-composite image of the central peak of South Ray with a bird's-eye-view image. **b**, Colour-composite and bird's-eye-view images of Tycho's central peak. **c**, Colour-composite and bird's-eye-view images of Tsiolkovsky's central peak. **d**, Colour-composite and bird's-eye-view images of a part of the Oriente (Inner Rook ring). All bluish areas in these composite images have modal abundances of plagioclase at nearly 98 vol.%. Bluish areas appear to be limited to the steep slopes where the subsurface layers might be exposed by landslides. **e**, Reflectance spectra of Jackson (J5 in Fig. 2), South Ray (S1), Tycho (Ty1), Tsiolkovsky (Ts1), and Oriente (O1). **f**, Absorption depths derived from the spectra in **e**. At Tycho, the PAN rocks are at the base of the central peak surrounding the relatively mafic-rich units (the yellow and green units in **b**). The morphology, mineralogical information, and slope angles observed in Tycho and Tsiolkovsky possibly suggest the impact melt origin of the darker unit found at the summit of each central peak. PAN rocks were found in South Ray, which is located near the Apollo 16 landing site. This is consistent with the fact that a large portion of the lunar samples with the highest plagioclase abundance were found among the Apollo 16 samples.

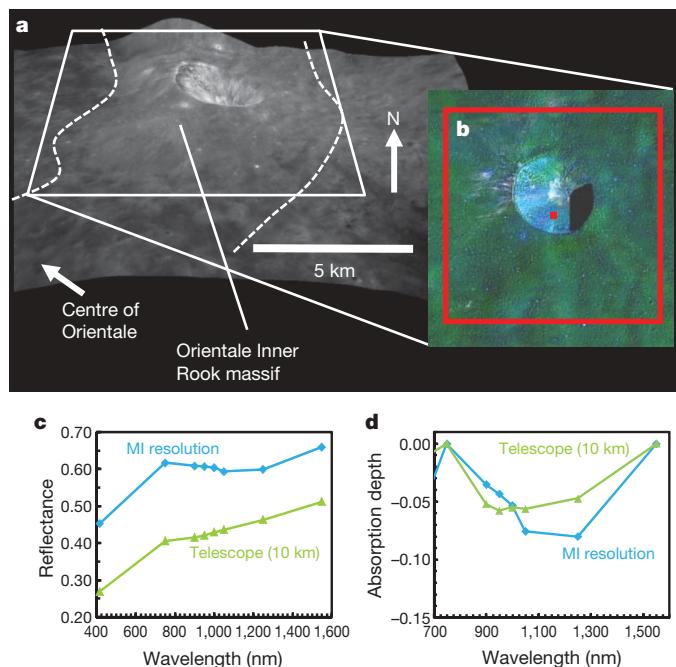


Figure 4 | Comparison of two different spatial resolutions at Oriente. **a**, Bird's-eye-view of the eastern portion of the Oriente Inner Rook (the exact location investigated in a previous study with an Earth-based telescope¹³ in which no clear plagioclase absorption band was found) produced from Multiband Imager (MI) 750-nm-band images superimposed on the DTM. **b**, Colour-composite image of area indicated in the bird's-eye-view image. **c**, Reflectance spectra presented as two different spatial resolutions. Both spectra use the same pixel as the centre pixel. Locations are indicated as red squares in **b**. **d**, Wavelengths versus absorption depths of the same spectra in **c** after the continuum removal. The depth of the Fe-bearing plagioclase absorption band in this outcrop decreases with decreasing spatial resolution. Shock deformation¹³ and space weathering²³ have been proposed as possible causes for the absence of plagioclase absorption.

We observed plagioclase abundance in the PAN rocks (~98 vol.%) that is much higher than previously estimated (82% to 92%)^{2,6,7}. If we assume the extreme case to be upper crust with the purest anorthositic composition, which is probably over-simplifying, then the lunar upper crust is estimated to be more enriched in Al_2O_3 (~35 wt%, because plagioclase in ferroan anorthosite has 36 ± 1 wt% aluminium content) than previously proposed ($\text{Al}_2\text{O}_3 = 24\text{--}32.2$ wt%)^{2,8,12}. This estimated value as an extreme case corresponds to an aluminium content of the bulk lunar crust that is 10% higher, assuming that roughly half of the crustal volume consists of the upper crust and given the aluminium content in ref. 7.

The global PAN rocks are probably formed by the crystallization and segregation of plagioclase inside a magma ocean of the Moon and thus support the existence of such a magma ocean. Previous crustal genesis models were constructed from earlier estimates of plagioclase abundance. Therefore, the PAN upper crust is a primary constraint that requires mechanisms to remove trapped liquid from gaps between plagioclase crystals because plagioclase has a wetting angle of ~45° (ref. 25) and removing the last several per cent of liquid would be very difficult. The mechanism is probably related to an upward force caused by buoyancy and recrystallization of plagioclase crystals²⁶. On the Earth, the purest anorthosite is produced by deformation²⁷. Therefore, considering the extremely high plagioclase abundance, similar to the Earth's purest anorthosite, PAN rocks might also be generated by deformation. The remaining liquid may have produced relatively mafic-rich anorthosite^{2,9} with a range of Fe/Mg ratios in mafic silicates.

METHODS SUMMARY

The spectral accuracy of the Multiband Imager instrument was confirmed by comparing spectra obtained by the Multiband Imager with those obtained by an Earth-based telescope²⁰. DTMs can be generated from Multiband Imager band sets that have an 11.2° maximum parallax. Usually, DTMs are generated by using nadir and most-slant bands in the visible spectrum. We confirmed the accuracy of the DTMs derived from the Multiband Imager to ± 150 m by comparing them with the DTMs generated by the Terrain Camera aboard SELENE and the Apollo topophotomap. The observed brightness depends on the local topography. In this paper, photometric correction using DTMs derived from the Multiband Imager has been applied to all images and spectra in Figs 2, 3 and 4. Absorption depths were derived after dividing each reflectance spectrum by its continuum (a line connecting the reflectance values in the log scale between two optimized wavelengths). The mineral mixing model is applied to all spectra after the continuum removal. See ref. 11 for more detail on the Multiband Imager calibration and correction procedures.

We selected and analysed 69 locations distributed all over the Moon, including 29 large young (Copernican or Eratosthenian) craters exceeding 40 km in diameter⁶, 15 previously studied²⁸ even younger smaller rayed craters, seven craters located on the margin of large basins¹³ and 18 bright craters chosen from USGS Clementine's 750-nm basemap, including two Apollo landing-site craters. Newly surfaced locations with minimum contamination from surrounding materials (Fig. 1) were used to avoid uncertainty caused by possible mixing with regolith. Locations with a low degree of optical maturity²⁹ (optical maturity index exceeding 0.6) are presented. However, Tsiolkovsky, which has an optical maturity index of 0.55, is an exception because its large diameter suggests that it has excavated a deep and possibly minimally disturbed crust.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 February; accepted 17 July 2009.

1. Taylor, S. R. *Planetary Science: A Lunar Perspective* (Lunar and Planetary Institute, 1982).
2. Warren, P. H. Lunar anorthosites and the magma-ocean plagioclase-flotation hypotheses: importance of FeO enrichment in the parent magma. *Am. Mineral.* **75**, 46–58 (1990).
3. Longhi, J. A new view of lunar ferroan anorthosites: postmagma ocean petrogenesis. *J. Geophys. Res.* **108**, doi:10.1029/2002JE001941 (2003).
4. Ohtake, M. et al. Performance and scientific objectives of the SELENE (KAGUYA) Multiband Imager. *Earth Planets Space* **60**, 257–264 (2008).
5. Kato, M., Sasaki, S., Tanaka, K., Iijima, Y. & Takizawa, Y. The Japanese lunar mission SELENE: science goals and present status. *Adv. Space Res.* **42**, 294–300 (2008).
6. Tompkins, S. & Pieters, C. M. Mineralogy of the lunar crust: results from Clementine. *Meteorit. Planet. Sci.* **34**, 25–41 (1999).

7. Wieczorek, M. A. et al. in *New Views of the Moon* (eds Jolliff, B. L. et al.) Vol. 60, Ch. 3, 221–343 (The Mineralogical Society of America, 2006).
8. Korotev, R. L., Jolliff, B. L., Zeigler, R. A., Gillis, J. J. & Haskin, L. A. Feldspathic lunar meteorites and their implications for compositional remote sensing of the lunar surface and the composition of the lunar crust. *Geochim. Cosmochim. Acta* **67**, 4895–4923 (2003).
9. Takeda, H. et al. Magnesian anorthosites and a deep crustal rock from the farside crust of the moon. *Earth Planet. Sci. Lett.* **247**, 171–184 (2006).
10. Clark, P. E. & Adler, I. Utilization of independent solar flux measurements to eliminate non-geochemical variation in X-ray fluorescence data. *Proc. Lunar Planet. Sci. Conf. IX*, 3029–3036 (1978).
11. Lawrence, D. J. et al. Thorium abundances on the lunar surface. *J. Geophys. Res.* **E 105**, 20307–20331 (1999).
12. Jolliff, B. L., Gillis, J. J., Haskin, L. A., Korotev, R. L. & Wieczorek, M. A. Major lunar crustal terranes: surface expressions and crust-mantle origins. *J. Geophys. Res.* **E 105**, 4197–4216 (2000).
13. Hawke, B. R. et al. Distribution and modes of occurrence of lunar anorthosite. *J. Geophys. Res.* **E 108**, doi:10.1029/2002JE001890 (2003).
14. Lucey, P. G. et al. in *New Views of the Moon* (eds Jolliff, B. L. et al.) Vol. 60, Ch. 2, 83–220 (The Mineralogical Society of America, 2006).
15. Adams, J. B. & Goulland, L. H. Plagioclase feldspars: visible and near infrared diffuse reflectance spectra as applied to remote sensing. *Proc. Lunar Planet. Sci. Conf. IX*, 2901–2909 (1978).
16. Pieters, C. M. The moon as a spectral calibration standard enabled by lunar samples: the Clementine example. *Workshop on New Views of the Moon II*, 8025–8026 (1999).
17. McEwen, A. et al. Summary of radiometric calibration and photometric normalizations steps for the Clementine UVVIS images. *Lunar Planet. Sci.* **XXIX**, 1466 (1998).
18. Spudis, P. D., Hawke, B. R. & Lucey, P. Composition of Orientale basin deposits and implications for the lunar basin-forming process. *J. Geophys. Res.* **B 89**, C197–C210 (1984).
19. Lucey, P. G. Model near-infrared optical constants of olivine and pyroxene as a function of iron content. *J. Geophys. Res.* **E 103**, 1703–1713 (1998).
20. Hawke, B. R., Lucey, P. G. & Bell, J. F. Spectral reflectance studies of Tycho crater: preliminary results. *Proc. Lunar Planet. Sci. Conf. XVII*, 999–1000 (1986).
21. McEwen, A. S. et al. Clementine observations of the Aristarchus region of the moon. *Science* **266**, 1858–1862 (1994).
22. Mouélic, S. L., Langevin, Y. & Erard, S. The distribution of olivine in the crater Aristarchus inferred from Clementine NIR data. *Geophys. Res. Lett.* **26**, 1195–1198 (1999).
23. Lucey, P. G. Radiative transfer model constraints on the shock state of remotely sensed lunar anorthosites. *Geophys. Res. Lett.* **29**, doi:10.1029/2001GL014655 (2002).
24. Cintala, M. J. & Grieve, R. A. F. Scaling impact melting and crater dimensions: implications for the lunar cratering record. *Meteorit. Planet. Sci.* **33**, 889–912 (1998).
25. Longhi, J. & Jurewicz, S. R. Plagioclase-melt wetting angle and textures: implications for anorthosites. *Proc. Lunar Planet. Sci. Conf. XXVI*, 859–860 (1995).
26. Phinney, W. C. & Morrison, D. A. Partition coefficient for calcic plagioclase: Implications for Archean anorthosites. *Geochim. Cosmochim. Acta* **54**, 1639–1654 (1990).
27. Lafrance, B., John, B. E. & Scoates, J. S. Syn-emplacement recrystallization and deformation microstructures in the Poe Mountain anorthosite, Wyoming. *Contrib. Mineral. Petrol.* **122**, 431–440 (1996).
28. Grier, J. A., McEwen, A. S., Lucey, P. G., Milazzo, M. & Strom, R. G. Optical maturity of ejecta from large rayed lunar craters. *J. Geophys. Res.* **E 106**, 32847–32862 (2001).
29. Lucey, P. G., Blewett, D. T., Taylor, G. J. & Hawke, B. R. Imaging of lunar surface maturity. *J. Geophys. Res.* **E 105**, 20377–20386 (2000).
30. Matsunaga, T. et al. Discoveries on the lithology of lunar crater central peaks by SELENE Spectral Profiler. *Geophys. Res. Lett.* **35**, doi:10.1029/2008GL035868 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank SELENE project team members Y. Takizawa, S. Sasaki, M. Kato and R. Nagashima. We also thank Fujitsu Limited engineers T. Maekawa, K. Tsubosaka, N. Tonoya, J. Inoue, N. Masuda and T. Nakashima. We are grateful to Mitsubishi Space Software Co. Ltd engineers M. Hashimoto, K. Torii, Y. Kurashina, A. Yoshizawa and S. Nakanotani. The long-term efforts by each of these teams were essential to our work. We are also grateful to P. G. Lucey for the optical constants of some minerals. The reviews by P. Warren and J. Longhi are much appreciated.

Author Contributions H.O., T.H., H.T., Y. Yamaguchi and T. Matsunaga suggested the original design of the Multiband Imager. M.O. finished the design and proposed the Multiband Imager observations. M.O., T. Matsunaga, J.H., H.O., J.T., T.S., N.H., R.N., H.D., S.M., S.K., Y. Yokota, T. Morota, C.H., Y.O., M.T., K.S., A.I. and N.A. developed the instrument and data processing system. T. Morota, C.H., M.T., M.O., T. Matsunaga, J.H., Y. Yokota, Y.O. and M.A. conducted the operation of the observation. M.O. and Y. Yokota conducted calibration and data analyses for this paper. M.O., T.A., H.M., H.T., Y. Yokota, N.H., R.N., T.H., J.H., T. Morota, K.K., T.S., K.S., T. Matsunaga, Y.O., S.S., A.Y. and C.M.P. contributed to writing the paper. All the authors, including M.S., H.A. and J.-L.J., discussed the results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.O. (ohtake.makiko@jaxa.jp).

METHODS

Data preparation. Multiband Imager data were calibrated⁴ by adjusting the data from the calibration standard target (Apollo 16 site) to the laboratory reflectance spectrum of an Apollo 16 soil sample¹⁶. The photometric function proposed by ref. 17 and DTMs generated by the Multiband Imager were used to convert the obtained data into the reflectance spectra in the standard viewing geometry. DTMs can be generated from Multiband Imager band sets that have an 11.2° maximum parallax. Usually, DTMs are generated by using nadir and most-slant bands in the visible spectrum. Instrumental errors of Multiband-Imager-derived signals are estimated as less than $\pm 1\%$. Multiband Imager images (or mosaics) from each location were selected for analysis, and pixel values of the images were adjusted to indicate the reflectance value according to the scaling factor that is indicated in the image header, which is usually required by remote sensing data. All images, including colour-composite images, are presented as 20 m per pixel. We used ENVI and IDL (<http://www.itvvis.com/ProductServices/ENVI.aspx>) for our data analyses.

Bird's-eye-view images. Bird's-eye-view images were produced from Multiband Imager 750-nm-band images superimposed on the DTM that was derived by the Multiband Imager. The topography was not exaggerated.

Colour-composite images. In this paper, red, green and blue were assigned to continuum-removed absorption depths of 950 nm, 1,050 nm and 1,250 nm to generate colour-composite images. Absorption depths were derived by dividing each reflectance spectrum by its continuum. A continuum was defined as a line connecting the reflectance values in the log-scale between two optimized wavelengths (750 nm or 900 nm, and 1,250 nm or 1,550 nm) selected for each spectrum.

Absolute reflectance spectra. All the reflectance spectra were given as an average of a 120 m \times 120 m area, which corresponds to 6 \times 6 pixels of each image, to remove spatial variation. If a small gap exists between the visible (415 nm, 750 nm, 900 nm, 950 nm and 1,000 nm) and near-infrared (1,000 nm, 1,050 nm, 1,250 nm and 1,550 nm) bands, the near-infrared bands are adjusted to visible bands using the correction factor (a factor for multiplication) for the 1,000-nm band. However, the difference between intensities at 1,000 nm obtained by near-infrared and visible sensors was generally small.

Absorption depths for each reflectance spectra. Absorption depths for each reflectance spectra were derived by procedures similar to that used to generate colour-composite images from each acquired absolute reflectance spectrum.

Absorption depths of the representative lunar minerals. The absorption depths of the lunar impact melt (77075) and the lunar minerals separated from Apollo samples of orthopyroxene (opx: 78235), clinopyroxene (cpx: 12063), olivine (ol: 72415), and plagioclase (pl: 15415) are presented for comparison.

Spectral reflectance data of these lunar minerals are from the RELAB database at Brown University (<http://www.planetary.brown.edu/rehab/>). Fractions (for example, 1/4) in the legend indicate that the absorption depth is multiplied by these fractions for comparison with plagioclase because the absorption depth of plagioclase is usually very shallow compared to that of mafic minerals.

Intimate mixing model. The modal abundance of the derived reflectance spectra was estimated by using intimate mixing models^{19,31}, in which the spectra of the target material can be reconstructed by mixtures of the spectra for end-member minerals as a function of their relative abundance. We used the end-member composition of plagioclase (FeO content 0.25 wt%), orthopyroxene (Ca₃Fe₄₀Mg₅₇), and clinopyroxene (Ca₃₁Fe₂₂Mg₄₇). The grain size for each end-member mineral was assumed to be the same and was adjusted to generate the observed absorption depth. We note that adding only a small amount (several per cent in volume) of mafic minerals to the plagioclase drastically changes and conceals the absorption of plagioclase, which is demonstrated in Fig. 2h and i. Olivine has an absorption band at 1,200 nm to 1,300 nm; however, the overall absorption depth originating from olivine is always strongest at around 1,050 nm in any composition³². Therefore, even though the absorption depth at 1,250 nm may be increased by the presence of olivine, there is no way to generate a greater absorption depth at 1,250 nm than observed at 1,050 nm by olivine. Thus, a greater absorption depth at 1,250 nm than at 1,050 nm or 1,000 nm or 950 nm is a basic criterion for identifying PAN rocks. The mineral mixing model that we used was an intimate mixture model and did not consider the spatial mixing of spectra. Modal abundances of plagioclase were estimated by using the absorption depth ratio between 1,000 nm and 1,250 nm as derived by the intimate mixing model. Estimated plagioclase modal abundances for these areas change only a little with the compositional change of the end-member minerals.

Evaluation of optical maturity. Newly surfaced locations that suffered minimum contamination with surrounding materials were used for the discussion in this paper to avoid the uncertainty caused by a possible mixing with regolith. Locations with a low degree of optical maturity²⁹ (optical maturity index exceeding 0.6) were selected and presented in Fig. 1 after the completion of the full data analysis described here. Tsiolkovsky, which has an optical maturity index of 0.55, was an exception because its large diameter indicates that it excavates a deep and possibly minimally disturbed crust.

31. Denevi, W., Lucey, P. G., Hochberg, E. J. & Steutel, D. Near-infrared optical constants of pyroxene as a function of iron and calcium content. *J. Geophys. Res.* **E 112**, doi:10.1029/2006JE002802 (2007).
32. Sunshine, J. M. & Pieters, C. M. Determining the composition of olivine from reflectance spectroscopy. *J. Geophys. Res.* **E 103**, 13675–13688 (1998).

Coherent optical pulse sequencer for quantum applications

Mahdi Hosseini¹, Ben M. Sparkes¹, Gabriel Hétet¹, Jevon J. Longdell^{1,2}, Ping Koy Lam¹ & Ben C. Buchler¹

The bandwidth and versatility of optical devices have revolutionized information technology systems and communication networks. Precise and arbitrary control of an optical field that preserves optical coherence is an important requisite for many proposed photonic technologies. For quantum information applications^{1,2}, a device that allows storage and on-demand retrieval of arbitrary quantum states of light would form an ideal quantum optical memory. Recently, significant progress has been made in implementing atomic quantum memories using electromagnetically induced transparency, photon echo spectroscopy, off-resonance Raman spectroscopy and other atom–light interaction processes. Single-photon^{3,4} and bright-optical-field^{5,6} storage with quantum states have both been successfully demonstrated. Here we present a coherent optical memory based on photon echoes induced through controlled reversible inhomogeneous broadening. Our scheme allows storage of multiple pulses of light within a chosen frequency bandwidth, and stored pulses can be recalled in arbitrary order with any chosen delay between each recalled pulse. Furthermore, pulses can be time-compressed, time-stretched or split into multiple smaller pulses and recalled in several pieces at chosen times. Although our experimental results are so far limited to classical light pulses, our technique should enable the construction of an optical random-access memory for time-bin quantum information, and have potential applications in quantum information processing.

Photon echo techniques show great promise as a form of quantum memory. So far there have been demonstrations of coherent pulse storage using techniques based on controlled reversible inhomogeneous broadening^{7–9} and atomic frequency combs¹⁰. The gradient echo memory (GEM) is a particularly promising photon echo system^{11–13} in which ensembles of two-level atoms are used as the storage medium. This scheme requires no auxiliary optical fields that can introduce noise and is, in principle, a 100% efficient multimode memory.

A two-level memory requires an optical transition with a long upper-state lifetime; otherwise, the storage time will be limited by the decay of the excited state. A long-lived state, however, implies a weakly interacting transition leading to low optical depth and absorption efficiency. This conundrum can be overcome by extending GEM to three-level systems¹⁴. As illustrated in Fig. 1a, a three-level atom can be addressed using a weak probe beam of amplitude ϵ_p , which we wish to store, and a strong coupling beam of amplitude ϵ_c with Rabi frequency Ω_c . The field ϵ_c mediates the coupling of the probe to a transition between two ground states. In the limit in which the one-photon detuning, Δ , is large in comparison with the upper-state decay rate, the weak probe field interacts with an effective two-level atom composed of the atomic ground states^{15,16}. The storage time of this system will be limited by the ground-state decoherence time, which for some systems can be many seconds¹⁷. The strength of the transition, on the other hand, is controlled by the coupling field.

A three-level GEM system thus isolates the optical depth from the storage time, which is a significant advantage over the two-level system. As will be shown, three-level GEM also allows us to recall stored information in any order.

The key to two- or three-level GEM is the application of an atomic frequency gradient, η , along the length of the storage medium (the z direction in Fig. 1). Depending on the atomic system, a linearly varying electric or magnetic field can be used to induce a Stark or Zeeman shift that varies in the z direction, as shown in Fig. 1b. In the most simple storage protocol, a probe field is absorbed by the frequency-shifted ensemble of atoms with effective linear density N , as shown in Fig. 1c. Owing to the frequency gradient in the ensemble, the Fourier components of the probe field are distributed linearly along on the z axis. The magnitude of the atomic polarization ($\hat{\sigma}_{12}$) in the z direction is therefore proportional to the Fourier spectrum of the optical field. At some time, τ , the frequency gradient is reversed, to $-\eta$. This reverses the evolution of the atomic dipoles. At time 2τ ,

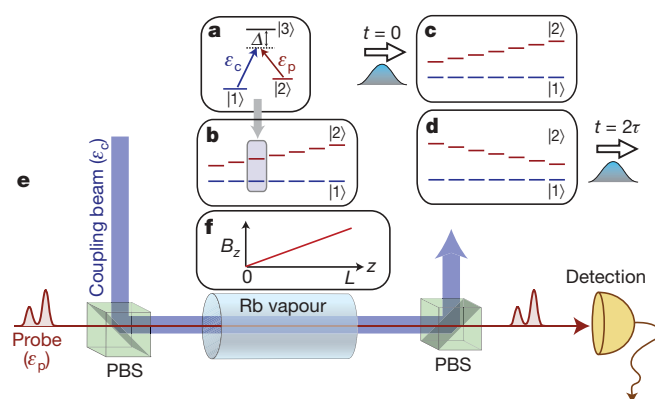


Figure 1 | GEM schematic. **a**, A three-level system can be addressed using a detuned coupling beam, ϵ_c . The ground states form an effective two-level system for the probe beam, ϵ_p . **b**, An ensemble of atoms with linearly varying frequency shift in the z direction. **c**, A pulse of light is stored in the frequency-shifted ensemble. **d**, After reversal of the frequency gradient at time τ , a photon echo emerges at time 2τ . **e**, The optical layout. Orthogonal, linearly polarized coupling and probe fields were sent through a polarizing beam splitter (PBS) into a warm Rb⁸⁷-enhanced cell with 1 torr of Kr buffer gas. The double-layer μ -metal shielded gas cell was surrounded by two variable-pitch coils that were used to apply magnetic field gradients in opposing directions. In our experiment the states $|1\rangle$, $|2\rangle$ and $|3\rangle$ as shown in **a** correspond to the $|5^2S_{1/2}, F=1\rangle$, $|5^2S_{1/2}, F=2\rangle$ and $|5^2P_{1/2}, F=2\rangle$ states of Rb⁸⁷. The two hyperfine ground states have a splitting of 6.8 GHz. Our probe beam was prepared using a fibre-coupled phase modulator driven at 6.8 GHz. A single sideband was selected using an optical cavity. The Raman resonance frequency varies linearly along the length of the cell as a result of the applied magnetic field gradient. **f**, The applied magnetic field, B_z .

¹ARC Centre of Excellence for Quantum-Atom Optics, Department of Quantum Science, The Australian National University, Canberra, Australian Capital Territory 0200, Australia.

²Physics Department, University of Otago, Dunedin 9016, New Zealand.

all the dipoles in the ensemble rephase and a photon echo is emitted in the forward direction, as shown in Fig. 1d. The atomic frequency gradient along the direction of propagation ensures that there is no re-absorption of light as the pulse is re-emitted, as none of the atoms will be resonant with the optical frequencies in the emerging pulse. This is in contrast to standard controlled reversible inhomogeneous broadening, in which recall of light in the forward direction is limited to 54% efficiency by re-absorption^{13,18}.

Our experiment is illustrated in Fig. 1e. An ensemble of three-level atoms, in our case a warm gas of Rb⁸⁷, was subjected to a strong coupling field detuned from resonance with $\Delta \approx 2$ GHz. The Raman absorption of the probe typically had a visibility of 85% and a width of 120 kHz. To create the atomic frequency gradient, we used a solenoid with variable winding pitch to create a linearly varying magnetic field, as shown in Fig. 1f. This gave a linearly varying Zeeman shift, meaning that the Raman absorption frequency within our ensemble varied linearly with z . For photon echo recall, a second variable-pitch solenoid with opposite current was used to invert η . The size of η determines the storage bandwidth of the memory. In our system, the magnetically broadened ensemble had Raman absorption widths of up to 1 MHz. The magnetic broadening decreased the effective optical depth such that the absorption was reduced to $\sim 60\%$. Free atoms have no linear Stark shift, so we had to use a magnetic field in our experiment. In other systems, for example rare-earth ions in solid state, electric fields can instead be used to shift the atomic levels¹¹.

To use our system as a coherent optical memory in its simplest form, a coupling beam and atomic frequency gradient were applied using the pattern shown in Fig. 2a to store a train of four pulses. It is not a requirement that ε_c be switched off during the storage phase,

but it is beneficial in practice because it eliminates spontaneous emission from the excited state. The experimental data are shown in Fig. 2c. The most striking feature of this result is that the shape of the input pulse train is reversed in time, as predicted previously¹². In this set-up, GEM is a first-in-last-out (FILO) memory.

A lossless, decoherence-free simulation (Methods Summary) of this experiment is shown in Fig. 2d. A train of four pulses enters the cell, which spans the length of the z axis, and is absorbed. The pulse train emerges in the forward direction symmetrically about the point of frequency-gradient switching. The model also shows that the pulse sequence has been reversed. The behaviour of our system, including this sequence reversal, is best understood using a Fourier decomposition of the optical and atomic fields in the spatial frequency (k) domain. As for two-level GEM¹², we find a normal mode $\hat{\psi}(t, k) = k\varepsilon_p(t, k)/\Omega_c + N\hat{\sigma}_{12}(t, k)/\Delta$ that propagates in the t - k plane according to

$$\left(\frac{\partial}{\partial t} - \eta(t)\frac{\partial}{\partial k} - i\frac{gN\Omega_c^2}{k\Delta^2}\right)\hat{\psi}(t, k) = 0 \quad (1)$$

Like the normal mode in electromagnetically induced transparency¹⁹ (EIT), $\hat{\psi}(t, k)$ is a combination of atomic polarization and optical field. In GEM, however, the normal mode is defined in the spatial Fourier domain. Equation (1) shows that the ‘speed’ of $\hat{\psi}(t, k)$ in the k direction is given by the frequency gradient, $\eta(t)$. The inset of Fig. 2c shows the evolution of $|\psi(t, k)|^2$ for the real-space data of Fig. 2d. The mode starts at $k = 0$ and evolves to higher k values at a rate determined by $\eta(t)$ until the frequency gradient is switched, leading to a reversal in propagation direction. The pulse is re-emitted when the mode returns to $k = 0$. A cross-section through $|\psi(t, k)|^2$ at any time

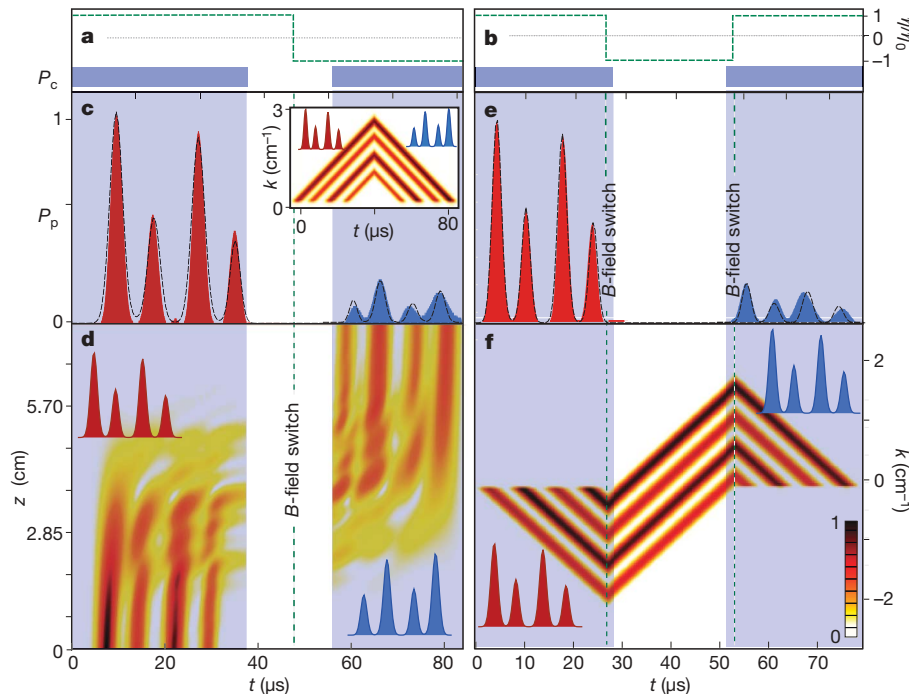


Figure 2 | First-in-last-out (FILO) and first-in-first-out (FIFO) memory. **a, b**, Switching scheme for FILO (**a**) and FIFO (**b**) storage showing the frequency gradient, η (dashed lines) normalized to the initial gradient η_0 , and the presence of the coupling field with power P_c (grey shading). **c**, Experimental observation of FILO storage showing P_p , the normalized power of the probe beam. Input pulses are shown in red and the photon echo (blue) shows order reversal. The frequency gradient was flipped at $t = 30 \mu\text{s}$. Dashed lines show a numerical simulation using the parameters $\Delta = 320\gamma$, $\eta L = 0.08\gamma$ and $\gamma_0 = 4$ kHz and an optical depth of $gNL/\gamma = 1.5$, where γ is the excited-state decay rate and g is the atom–light coupling strength for the $|1\rangle \leftrightarrow |3\rangle$ transition. The output echo and simulation are magnified by a

factor of ten. Inset, dynamics of $|\psi(t, k)|^2$ for this storage scheme (colour scale in **f**). **d**, A decoherence-free numerical simulation showing P_p (colour scale in **f**) in the t - z plane for FILO memory. The input pulse sequence (red) is reversed at the output (blue). **e**, Experimental observation of FIFO retrieval: red, input pulses; blue, photon echo showing order preservation. The dashed line shows numerical modelling using the same parameters as in **c**, except with $\gamma_0 = 3$ kHz. The output echo and simulation are magnified by a factor of ten. **f**, A decoherence-free numerical simulation of $|\psi(t, k)|^2$ for FIFO storage. In this case, the coupling beam is off when the normal mode crosses $k = 0$.

is proportional to the temporal profile of the input optical-field intensity. This is because, as described above, the optical field is stored as a spatial Fourier transform in the z -axis, so a second Fourier transform into k -space returns the original pulse shape. In this picture, the reason for the pulse sequence reversal is clear: the last pulse to enter the system returns to $k = 0$ first and is thus re-emitted first. Including the ground-state decoherence rate, γ_0 , and N as free parameters in our numerical model, we were able to fit the data in Fig. 2c with excellent agreement, as shown by the dashed lines.

A natural question is whether it is possible to avoid reversing the pulse shape. This would require that the normal mode return to $k = 0$ travelling in the positive k direction. The last pulse in would then be the last pulse out. With a two-level system, this would seem impossible: after reversing the frequency gradient there is no way to suppress the emission when the normal mode returns to $k = 0$. With a three-level system, however, we are free to turn off the coupling beam. In this case, although the dipoles will rephase when the normal mode reaches $k = 0$, no light can be emitted. This is seen in Fig. 2f, which shows $|\psi(t, k)|^2$ for the switching scheme in Fig. 2b. With the coupling beam switched off, the normal mode passes straight through $k = 0$ to negative k values. We can then reverse the frequency gradient again, to obtain a normal mode travelling in the positive k direction. With the coupling field switched back on, the normal mode is converted into a photon echo at $k = 0$ without pulse shape reversal. This is demonstrated experimentally in Fig. 2e. In this way, we can construct a first-in-first-out (FIFO) memory. As in the case of FILO memory, our numerical model (dashed curve) shows excellent agreement with the data.

Combining the FIFO and FILO techniques, our system can be thought of as a k -space 'conveyor belt' for the stored light pulses. The normal mode can be moved back and forth along the k -axis by controlling the frequency gradient, η . Furthermore, we are able to push pulses off the conveyor belt whenever they pass through $k = 0$, by turning on the coupling beam. In this way, we are able to construct a system that can recall the pulses in any order we choose. A decoherence-free model of this on-demand retrieval is shown in Fig. 3. This figure demonstrates not just the arbitrary recovery of pulses, but also methods of manipulating the pulses during storage.

We start with seven pulses of varying intensities. The pulses are first read into the memory with $\eta = \eta_0$ and are then held in a steady state with $\eta = 0$. Reversing the motion in k -space with $\eta = -\eta_0$ allows us to couple out some pulses in reverse order as they pass through $k = 0$. By timing the coupling field correctly, we recover pulses 3 and 4 (Fig. 3b, c) at $t = 15\tau_p$, where τ_p is the width of a single pulse. Next we hold the normal mode for some time at negative k values, before switching to a positive velocity along the k -axis. By again timing the coupling beam

correctly, we recover pulses 5 and 6 without reversal at $t = 27\tau_p$. The remaining pulses move to positive k values. On the next pass through $k = 0$, we reduce the power of the coupling beam by a factor of two. This allows us to couple out half the power of pulses 1 and 2 in reverse order at $t = 34\tau_p$. This system amounts to a beam splitter with a variable time delay at one port, as we are free to recover the rest of these pulses at a later time. We then switch back to a positive slope, but this time with a higher frequency gradient, $\eta = 4\eta_0$. This causes pulse compression because increasing η widens the range of frequencies covered by the atomic ensemble and a wider Fourier width leads to shorter pulses. This can be understood intuitively in k -space, as the normal mode moves faster through $k = 0$, leading to faster pulse recovery. In this way, half of the remainders of pulses 1 and 2 are compressed and released from the memory in their original order at $t = 37\tau_p$. In the last stage, we reduce η to achieve pulse stretching. The expanded remains of pulses 1 and 2 are thus released from the memory in reverse order at $t = 45\tau_p$. The last pulse (pulse 7) is left in the atomic medium.

Experimental demonstrations of the various recall techniques discussed above are shown in Fig. 4. There we show pulse re-ordering (Fig. 4a), splitting of a pair of pulses over two recall events (Fig. 4b) and pulse-width modification (Fig. 4c), using different frequency gradients. The model shown in Fig. 3 relied on reduced coupling-beam power to induce pulse splitting. In our experiment, we find that we can split pulses with constant values of ε_c . This is due to the low optical depth in our system, which limits both the writing and read-out stages of the photon echo. Inefficient recall allows us simply to read out twice without changing ε_c , as predicted previously¹³. Numerical modelling (Fig. 4, dashed lines) again shows excellent agreement with our experimental data.

Although GEM can theoretically reach 100% efficiency, the multimode results we have presented so far have efficiencies of $\sim 5\%$. This compares favourably with the atomic frequency comb¹⁰ multimode memory with $<0.5\%$ recall. For a single temporal mode and short storage times, our system can achieve a recall efficiency of 41%, as shown in Fig. 4d. For a single-mode storage time equal to the pulse duration, we achieve 31% recall efficiency. This is comparable to the state of the art in single-mode EIT systems, which have 42% recall efficiency²⁰. The optical depth, which limits our efficiency, could be improved by using optical pumping to increase the number of effective atoms. Ultimately, however, warm atomic ensembles are restricted by atomic motion and collisional broadening. High-efficiency GEM for quantum information systems will most likely require more advanced atomic systems. Cold clouds of alkali atoms, for example, have superior optical depths and longer coherence

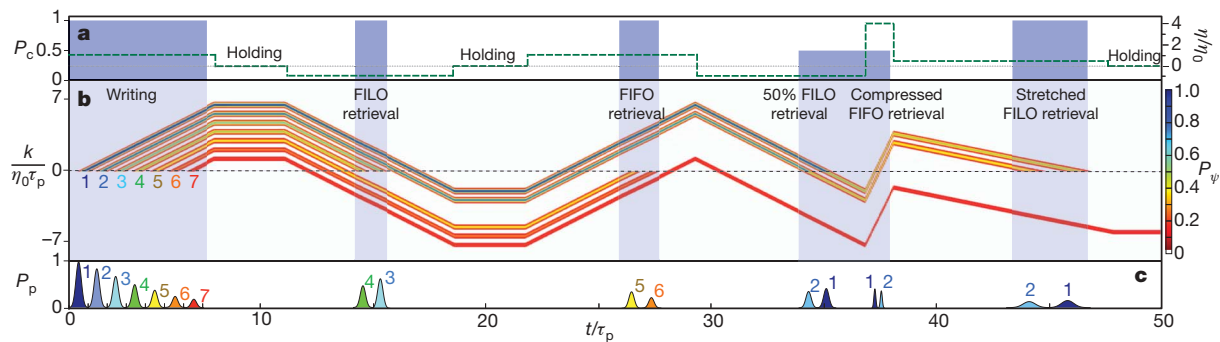


Figure 3 | The coherent optical pulse sequencer. **a**, Switching algorithm for the frequency gradient, η (dashed line), and optical coupling field, P_c (shading). **b**, Evolution of the k -space normal-mode power, $P_\psi = |\psi(t, k)|^2$, for seven input pulses. **c**, Temporal profile for the input and recalled pulses. Writing of the pulses occurs with the frequency gradient and coupling field switched on. The normal modes can be kept in a holding pattern by turning both fields off. Optical pulses are retrieved by turning on the coupling field when the normal mode crosses $k = 0$. FILO retrieval of pulses (4, 3) is

achieved using a negative frequency gradient. FIFO retrieval of pulses (5, 6) is achieved using a positive frequency gradient. Partial retrieval of pulses (2, 1) is achieved with reduced coupling power. Variation of the frequency gradient is used to stretch or compress pulses 1 and 2 in further recall events. Parameters used are $\eta L = 48\gamma$, $\Omega_c = 35\gamma$, $g = 3\gamma$, $P_{p, \max} = 2.5 \times 10^{-7} P_c$, $\Delta = 1,000\gamma$, $\gamma_0 = 0$ and $\gamma = 1$, and the optical depth is $gNL/\gamma = 600$. All quantities plotted are expressed in normalized units.

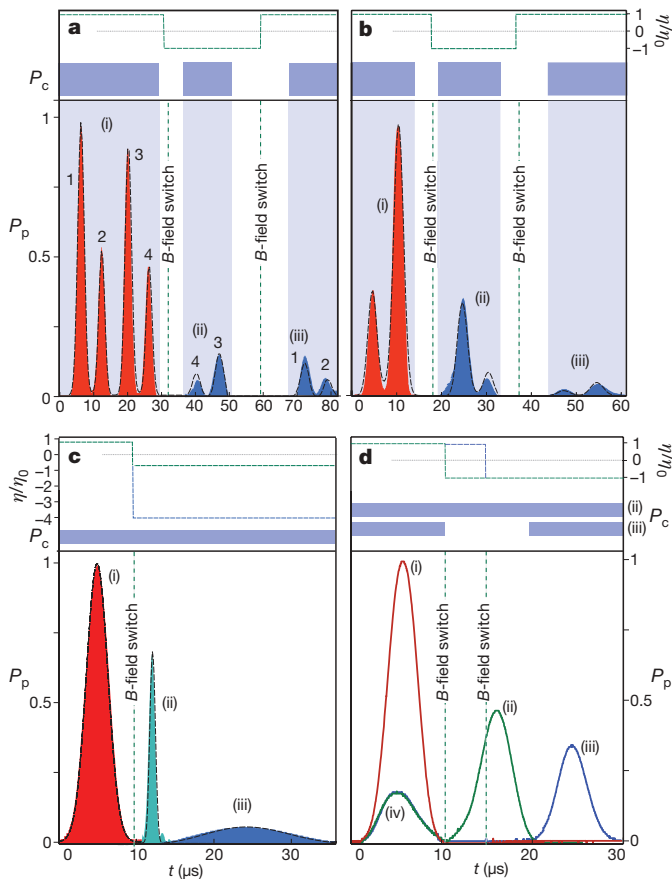


Figure 4 | Flexible pulse recall. The switching patterns are shown at the top of each panel, with the normalized frequency gradient, η , denoted by dashed lines and the presence of the control beam, P_c , indicated by grey shading. **a**, Pulse re-ordering: (i) four input pulses are written into the memory; (ii) after the first frequency-gradient switch, FILO retrieval of pulses 4 and 3 is observed; (iii) a second frequency-gradient switch produces FIFO retrieval of pulses 1 and 2. The output echo and simulation are magnified by a factor of ten. **b**, Pulse splitting: (i) two input pulses are written into the memory; (ii) partial FILO retrieval of the input follows immediately; (iii) a second partial FIFO retrieval of the input follows at later time. The output echo and simulation are magnified by a factor of ten. **c**, Compression and expansion: separate experiments show (i) time-compressed retrieval with $\eta = -4\eta_0$; (ii) time-stretched retrieval with $\eta = -0.6\eta_0$. Data shown without magnification. **d**, High-efficiency single-pulse storage: (i) input pulse; (ii) 42% recall efficiency; (iii) 31% recall efficiency; (iv) leakage of 20% of the input through the cell, owing to limited optical depth. The dashed lines show numerical simulations of Gaussian pulses with $\gamma_0 = 5$ kHz and $\eta L = 0.06\gamma$ (**a**), $\gamma_0 = 1$ kHz and $\eta L = 0.08\gamma$ (**b**), $\gamma_0 = 9$ kHz and $\eta L = 0.07\gamma$ (**c**). In all cases, the optical depth is $gNL/\gamma = 1.5$ and the detuning is $\Delta = 320\gamma$.

times. Cryogenic rare-earth-doped crystals are another option. These systems can have high optical depths and ground-state coherence times of many seconds¹⁷, and GEM can take advantage of the Stark shift for the atomic frequency gradient^{8,11}.

The GEM system has a number of possible applications in quantum science. Although our experiments were done with relatively bright pulses, the analysis is valid in the single-photon regime. Our system can not only store single photons and recall them in any order, but also makes it possible to control the bandwidth of the single photons by recalling using different frequency gradients. Recalling using a reduced gradient allows for narrow-bandwidth single photons, whereas recalling using an increased gradient will localize the photons temporally. Control of single-photon bandwidth is important for high-efficiency coupling to atomic resonances or optical cavities. Three-level GEM could also be applied in quantum repeaters that rely on multimode quantum memory²¹. The ability of our system to

recall information in any order may make an interesting addition to such repeater schemes. Moreover, in some quantum information protocols, temporal separation is used to distinguish ‘time-bin’ qubits^{22,23}. Such qubits could be stored using our pulse sequencer and then recalled in any order, effectively creating a random-access memory for time-bin quantum information. With the expected improvements in efficiency that will come from the use of more suitable atomic ensembles, three-level GEM has the potential to play an important part in many future quantum information systems.

METHODS SUMMARY

Simulations of our system were performed by solving the Heisenberg–Langevin equations in the weak probe limit¹⁴:

$$\frac{\partial}{\partial t} \hat{\sigma}_{13} = -(\gamma + \gamma_0/2 + i\Delta) \hat{\sigma}_{13} + ig\hat{e}_p + i\Omega_c \hat{\sigma}_{12} + \hat{\mathcal{F}}_{13}$$

$$\frac{\partial}{\partial t} \hat{\sigma}_{12} = -(\gamma_0 + i\eta(t)z) \hat{\sigma}_{12} + i\Omega_c^* \hat{\sigma}_{13} + \hat{\mathcal{F}}_{12}$$

$$\frac{\partial}{\partial z} \hat{e}_p = iN\hat{\sigma}_{13}$$

Here $\hat{\sigma}_{ij}$ is the atomic coherence of the transition between states $|i\rangle$ and $|j\rangle$ and an asterisk denotes complex conjugation. In the weak probe limit, where the equations are linear, the Langevin noise operators, $\hat{\mathcal{F}}_{ij}$, do not contribute noise beyond that required to preserve the canonical commutation relations^{16,24}. We can then ignore the $\hat{\mathcal{F}}_{ij}$ terms and solve the equations numerically using ordinary numbers (c-numbers) for the operators²⁵.

Received 24 June; accepted 21 July 2009.

- Duan, L. M., Lukin, M. D., Cirac, J. I. & Zoller, P. Long-distance quantum communication with atomic ensembles and linear optics. *Nature* **414**, 413–418 (2001).
- Zoller, P. *et al.* Quantum information processing and communication. *Eur. Phys. J. D* **36**, 203–228 (2005).
- Eisaman, M. D. *et al.* Electromagnetically induced transparency with tunable single-photon pulses. *Nature* **438**, 837–841 (2005).
- Chaneli re, T. *et al.* Storage and retrieval of single photons transmitted between remote quantum memories. *Nature* **438**, 833–836 (2005).
- Honda, K. *et al.* Storage and retrieval of a squeezed vacuum. *Phys. Rev. Lett.* **100**, 093601 (2008).
- Appel, J., Figueroa, E., Korystov, D., Lobino, M. & Lvovsky, A. I. Quantum memory for squeezed light. *Phys. Rev. Lett.* **100**, 093602 (2008).
- Nilsson, M. & Kr  ll, S. Solid state quantum memory using complete absorption and reemission of photons by tailored and externally controlled inhomogeneous absorption profiles. *Opt. Commun.* **247**, 393–403 (2005).
- Alexander, A. L., Longdell, J. J., Sellars, M. J. & Manson, N. B. Photon echoes produced by switching electric fields. *Phys. Rev. Lett.* **96**, 043602 (2006).
- Alexander, A. L., Longdell, J. J., Sellars, M. J. & Manson, N. B. Coherent information storage with photon echoes produced by switching electric fields. *J. Lumin.* **127**, 94–97 (2007).
- de Riedmatten, H. & Afzelius, M. Staudt, M. U., Simon, C. & Gisin, N. A solid-state light–matter interface at the single-photon level. *Nature* **456**, 773–777 (2008).
- H  t  t, G., Longdell, J. J., Alexander, A. L., Lam, P. K. & Sellars, M. J. Electro-optic quantum memory for light using two-level atoms. *Phys. Rev. Lett.* **100**, 023601 (2008).
- H  t  t, G., Longdell, J. J., Sellars, M. J., Lam, P. K. & Buchler, B. C. Multimodal properties and dynamics of gradient echo quantum memory. *Phys. Rev. Lett.* **101**, 203601 (2008).
- Longdell, J. J., H  t  t, G., Lam, P. K. & Sellars, M. J. Analytic treatment of controlled reversible inhomogeneous broadening quantum memories for light using two-level atoms. *Phys. Rev. A* **78**, 032337 (2008).
- H  t  t, G. *et al.* Photon echoes generated by reversing magnetic field gradients in a rubidium vapor. *Opt. Lett.* **33**, 2323–2325 (2008).
- Raymer, M. G. & Mostowski, J. Stimulated Raman scattering: unified treatment of spontaneous initiation and spatial propagation. *Phys. Rev. A* **24**, 1980–1993 (1981).
- Gorshkov, A. V., Andr  , A., Lukin, M. D. & S  rensen, A. S. Photon storage in Λ -type optically dense atomic media. I. Cavity model. *Phys. Rev. A* **76**, 033804 (2007).
- Fraval, E., Sellars, M. J. & Longdell, J. J. Dynamic decoherence control of a solid-state nuclear-quadrupole qubit. *Phys. Rev. Lett.* **95**, 030506 (2005).
- Sangouard, N., Simon, C., Afzelius, M. & Gisin, N. Analysis of a quantum memory for photons based on controlled reversible inhomogeneous broadening. *Phys. Rev. A* **75**, 032327 (2007).
- Fleischhauer, M. & Lukin, M. D. Dark-state polaritons in electromagnetically induced transparency. *Phys. Rev. Lett.* **84**, 5094–5097 (2000).
- Novikova, I., Phillips, N. B. & Gorshkov, A. V. Optimal light storage with full pulse-shape control. *Phys. Rev. A* **78**, 021802(R) (2008).
- Simon, C. *et al.* Quantum repeaters with photon pair sources and multimode memories. *Phys. Rev. Lett.* **98**, 190503 (2007).

22. Brendel, J., Gisin, N., Tittel, W. & Zbinden, H. Pulsed energy-time entangled twin-photon source for quantum communication. *Phys. Rev. Lett.* **82**, 2594–2597 (1999).
23. Gisin, N., Moiseev, S. A. & Simon, C. Storage and retrieval of time-bin qubits with photon-echo-based quantum memories. *Phys. Rev. A* **76**, 014302 (2007).
24. Hétet, G., Peng, A., Johnsson, M. T., Hope, J. J. & Lam, P. K. Characterization of electromagnetically-induced-transparency-based continuous-variable quantum memories. *Phys. Rev. A* **77**, 012323 (2008).
25. Drummond, P. D. & Raymer, M. G. Stimulated Raman scattering: unified treatment of spontaneous initiation and spatial propagation. *Phys. Rev. A* **44**, 2072–2085 (1991).

Acknowledgements We thank S. Bell for the solenoid design code and J. Close, C. Savage and P. Drummond for suggestions. This work was supported by the Australian Research Council.

Author Contributions The experiments were designed by G.H., J.J.L., B.C.B., M.H. and P.K.L., built by M.H., B.M.S. and G.H., modelled by M.H. with assistance from G.H. and carried out by M.H. with assistance and supervision from B.C.B. and P.K.L. The manuscript was prepared by B.C.B., M.H. and P.K.L. and edited by all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to B.C.B. (ben.buchler@anu.edu.au).

LETTERS

Stable single-unit-cell nanosheets of zeolite MFI as active and long-lived catalysts

Minkee Choi^{1*}, Kyungsu Na^{1*}, Jeongnam Kim^{1,2}, Yasuhiro Sakamoto^{5,6}, Osamu Terasaki^{4,6} & Ryong Ryoo^{1,2,3}

Zeolites—microporous crystalline aluminosilicates—are widely used in petrochemistry and fine-chemical synthesis^{1–3} because strong acid sites within their uniform micropores enable size- and shape-selective catalysis. But the very presence of the micropores, with aperture diameters below 1 nm, often goes hand-in-hand with diffusion limitations^{3–5} that adversely affect catalytic activity. The problem can be overcome by reducing the thickness of the zeolite crystals, which reduces diffusion path lengths and thus improves molecular diffusion^{4,5}. This has been realized by synthesizing zeolite nanocrystals⁶, by exfoliating layered zeolites^{7–9}, and by introducing mesopores in the microporous material through templating strategies^{10–17} or demetallation processes^{18–22}. But except for the exfoliation, none of these strategies has produced ‘ultrathin’ zeolites with thicknesses below 5 nm. Here we show that appropriately designed bifunctional surfactants can direct the formation of zeolite structures on the mesoporous and microporous length scales simultaneously and thus yield MFI (ZSM-5, one of the most important catalysts in the petrochemical industry) zeolite nanosheets that are only 2 nm thick, which corresponds to the *b*-axis dimension of a single MFI unit cell. The large number of acid sites on the external surface of these zeolites renders them highly active for the catalytic conversion of large organic molecules, and the reduced crystal thickness facilitates diffusion and thereby dramatically suppresses catalyst deactivation through coke deposition during methanol-to-gasoline conversion. We expect that our synthesis approach could be applied to other zeolites to improve their performance in a range of important catalytic applications.

In principle, zeolites will exhibit maximized molecular diffusion if the thickness of the crystal is reduced to the single unit cell dimension. Isolated zeolite unit cells (zero-dimensional crystal structure), nanowires (one-dimensional) and nanosheets (two-dimensional) would be obtained when confining the crystal thickness to the dimension of a single unit cell along three, two and one spatial dimensions, respectively. Of these three types of ultrathin zeolites, nanowires and nanosheets would be preferred in heterogeneous catalysis because of their ease of handling (they are collectable by filtration).

Although such ultrathin zeolites are easy to imagine, their actual synthesis is extremely difficult. This is because zeolite crystallization, like other crystallization processes, is accompanied by Ostwald ripening. Ostwald ripening is a naturally and spontaneously occurring process that minimizes the surface free energy of crystals, resulting in dissolution of smaller crystals and further growth of larger crystals. This thermodynamically controlled phenomenon becomes more significant as the crystal size decreases and therefore renders the synthesis of ultrathin zeolites extremely challenging. Although careful optimization of crystallization conditions⁶, or use of solid templates^{10–13} or organosilanes^{14–16} during synthesis resulted in

zeolites with significantly reduced framework thicknesses, thicknesses were typically still between 5 and 100 nm. Only the layer-by-layer exfoliation of a MWW (MCM-22) zeolite crystal^{7–9} yielded zeolites with ultrathin frameworks less than 5 nm thick.

We approached the synthesis of MFI nanosheets by designing a di-quaternary ammonium-type surfactant, $C_{22}H_{45}-N^+(CH_3)_2-C_6H_{12}-N^+(CH_3)_2-C_6H_{13}$ (designated C_{22-6-6} hereafter)²³. The surfactant was composed of a long-chain alkyl group (C22) and two quaternary ammonium groups spaced by a C6 alkyl linkage (see Supplementary Fig. 1 for the three-dimensional molecular structure). The diammonium head group acted as an effective structure-directing agent for the MFI zeolite, while the hydrophobic interaction between the long-chain tails induced the formation of mesoscale micellar structure. With the surfactant, an ultrathin zeolite framework was formed at the hydrophilic part of the micelles while the hydrophobic tail restricted the excessive growth of zeolites. It is noteworthy that ordinary surfactants, with a single quaternary ammonium group, failed to function as an effective structure-directing agent for zeolite (generating amorphous MCM-41-type silicas)^{24,25}.

MFI zeolites with Si/Al ratio of 30 to ∞ were crystallized by using the diammonium surfactant as a structure-directing agent. In a typical synthesis condition (Methods), the zeolite was obtained as multilamellar stacking of MFI nanosheets that were three-dimensionally intergrown (Fig. 1a). The overall thickness of the lamellar stacking was normally 20–40 nm. High-resolution transmission electron microscope (TEM) investigation of the cross-section (Fig. 1c) revealed that the stacking was composed of alternating layers of 2.0-nm-thick MFI zeolite framework and 2.8-nm-thick surfactant micelles. The zeolite layer was composed of three pentasil sheets, which corresponded to a single unit cell dimension along the *b*-axis ($b = 1.9738$ nm). TEM investigation and electron diffraction on the layer surface identified it as the (010) surface of the MFI framework (Fig. 1b). The short arc in the electron diffraction pattern (Fig. 1b) indicated that each zeolite layer possessed high structural correlation in the *a*–*c* plane orientation with a minor deviation. Only the *h*0*l* reflections were sufficiently sharp for indexing in the powder X-ray diffraction pattern (Fig. 1d), confirming that the zeolite layer possessed large coherent domains characterized by wide *a*–*c* planes while the framework thickness along the *b*-axis was extremely small. Elemental analysis revealed that the surfactant content was about 45 wt% of the as-synthesized product. The surfactant content could be decreased to 19 wt% (that is, a SiO₂/surfactant molar ratio of 37) by extraction with an HCl/ethanol solution. The non-extractable content is supposed to be the amount entrapped in the zeolite micropores after the surfactant has acted as a structure-directing agent. The extractable portion is attributed to the surfactant molecules that are located in the surfactant micelle as ‘dummy’ filler. On the basis of the elemental analysis and

¹Center for Functional Nanomaterials, Department of Chemistry, KAIST, ²Graduate School of Nanoscience and Technology (WCU), KAIST, ³KAIST Institute for the NanoCentury, ⁴Graduate School of EEWS (WCU), KAIST, Daejeon 305-701, Korea. ⁵Structural Chemistry, Arrhenius Laboratory, Stockholm University, 10691 Stockholm, Sweden. ⁶Nanoscience and Nanotechnology Research Center, Osaka Prefecture University, Sakai 599-8570, Japan.

*These authors contributed equally to this work.

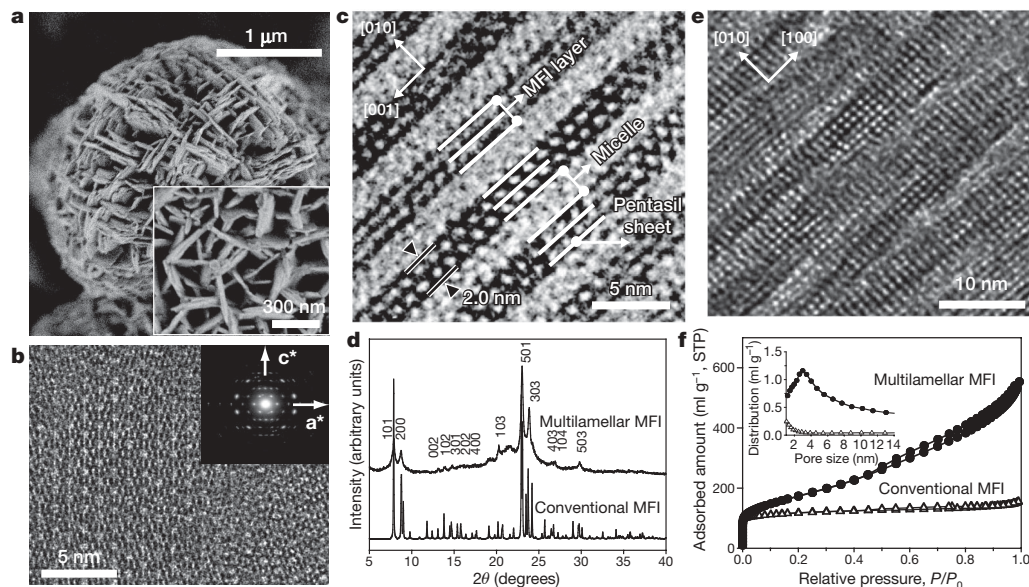


Figure 1 | MFI nanosheets with a multilamellar structure. **a–d**, As-synthesized sample; **e** and **f**, calcined sample. **a**, SEM image showing that the MFI zeolite has a plate-like morphology that is composed of three-dimensionally intergrown nanosheets. **b**, TEM and electron diffraction on the wide plane of the plate ([010] incidence of MFI). **c**, TEM cross-section of the plate revealing that the each plate is composed of lamellar stacking of alternating layers of MFI (2 nm) and surfactant micelle (2.8 nm). The MFI layer is composed of three pentasil sheets, corresponding to the thickness of a single unit cell dimension along the *b*-axis of $b = 1.9738$ nm. **d**, Powder

X-ray diffraction pattern indicating that only $h0l$ reflections are sufficiently sharp for indexing. The result confirms that the zeolite possesses wide *a*–*c* planes having large coherent domains, while the framework thickness along the *b*-axis is extremely small. **e**, TEM image of calcined sample showing that calcination leads to partial condensation between MFI layers, while the interlayer space (mesoporosity) is still mainly intact. **f**, N_2 adsorption–desorption isotherm, also confirming the highly mesoporous structure of the calcined sample. BET area = 520 m² g^{−1}. STP, standard temperature and pressure.

forementioned TEM investigations, we propose that the material is composed of MFI layers wherein the surfactant molecules are aligned along the straight micropores of the MFI framework (Fig. 2a).

Because the surfactant layers provide interlamellar support (Fig. 2b), surfactant removal was expected to lead to the complete condensation of the MFI layers. However, the calcination actually led to a partial condensation only (Fig. 1e). The calcined product was highly mesoporous, although the mesopore size distribution was rather broad

owing to the irregular distortion of zeolite layers (Fig. 1f). The calcined sample still exhibited a markedly enhanced Brunauer–Emmett–Teller (BET) area (520 m² g^{−1}) compared to conventional MFI zeolite (420 m² g^{−1}). The retained mesoporosity can be explained as follows. First, as indicated by the scanning electron microscope (SEM) image (Fig. 1a), there were a large number of crystal intergrowths. The intergrown crystals could act as a ‘pillar’ supporting each other, preventing complete collapse of the mesoporous structure. Second, there were

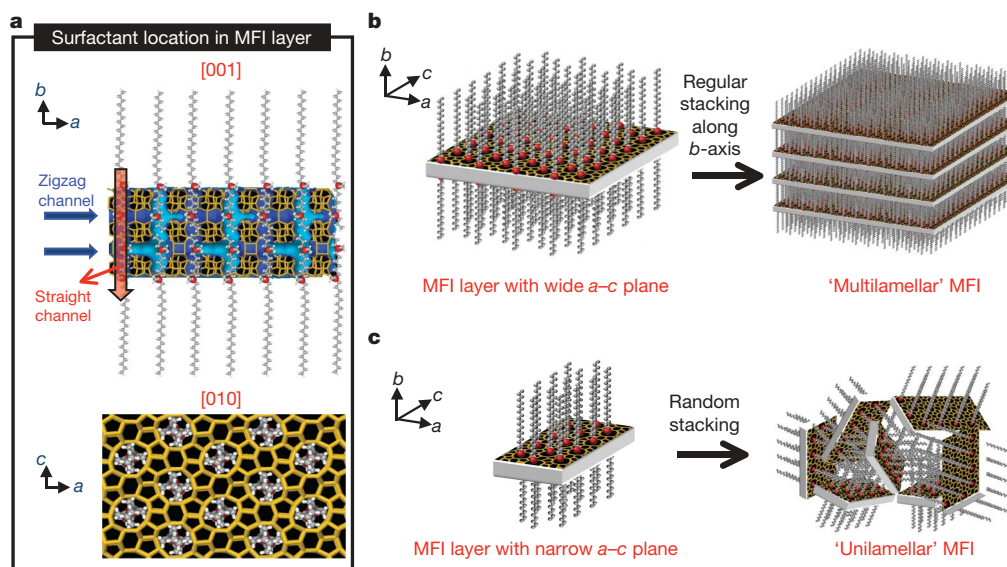


Figure 2 | Crystallization of MFI nanosheets. **a**, Proposed structure model for the single MFI nanosheet. Surfactant molecules are aligned along the straight channel of MFI framework. Two quaternary ammonium groups (indicated as a red sphere) are located at the channel intersections; one is

inside the framework, and the other is at the pore mouth of the external surface. Many MFI nanosheets form either multilamellar stacking along the *b*-axis (**b**), or a random assembly of unilamellar structure (**c**).

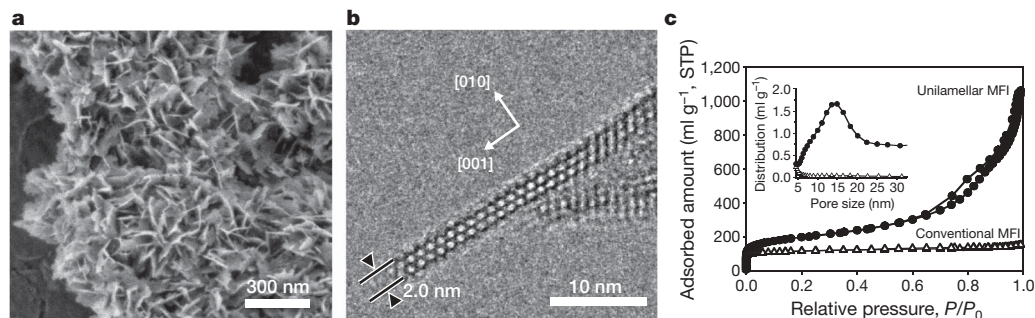


Figure 3 | MFI nanosheets with a unilamellar structure. **a**, SEM image showing that the MFI zeolite is synthesized in a flake-like morphology. **b**, TEM image of the cross-section of the flake, revealing that each flake is

composed of a discrete MFI layer having unit cell thickness along the *b*-axis of $b = 1.9738$ nm. **c**, N_2 adsorption-desorption isotherm confirming the highly mesoporous structure of the calcined sample. BET area = $710 \text{ m}^2 \text{ g}^{-1}$.

slight deviations of the crystal orientation in the *a*-*c* plane, and this mismatch between the position of silanol groups on each MFI layers prevented the complete condensation of MFI layers.

In addition to the multilamellar form, MFI zeolite could also be synthesized in the form of unilamellar nanosheets (Fig. 3a) by reducing the concentration of Na^+ in the synthesis mixture (Methods). TEM images (Fig. 3b) revealed that the material was composed of a single MFI layer (that is, three pentasil sheets) having a very narrow *a*-*c* plane; this material can thus be considered to be composed of essentially the same building blocks as the multilamellar form but without long-range stacking along the *b*-axis (Fig. 2c). The ability to produce these different forms indicates that crystal growth in the *a*-*c* plane and layer stacking along the *b*-axis are significantly affected by the concentration of Na^+ in the synthesis mixture. The unilamellar zeolite exhibited a significantly increased surface area ($710 \text{ m}^2 \text{ g}^{-1}$), compared to its multilamellar counterpart ($520 \text{ m}^2 \text{ g}^{-1}$) (Fig. 3c).

The catalytic performance of the MFI nanosheets was investigated using large organic molecules so that diffusion of the reactant molecules constrains the reaction (Methods)²⁶. As expected, the catalytic activities (per weight of catalyst) of the MFI nanosheets were much higher than those of conventional MFI zeolite (see Table 1). These enhanced catalytic activities can be attributed to a large number of acid sites located at the mesopore surface (that is, on the external surface of the zeolite layer) of MFI nanosheets, with the unilamellar MFI generally exhibiting higher activities owing to its larger external surface area after calcination.

Another remarkable feature of the MFI nanosheets is their increased catalyst lifetime, which manifested itself when we investigated the catalytic properties of MFI zeolites in methanol-to-gasoline conversion. Owing to methanol's small size, there was no significant difference in the initial catalytic activity between the ultrathin and the conventional MFI zeolite. With time on-stream, however, the MFI nanosheets were deactivated far more slowly than the conventional MFI (Fig. 4). To determine why, we monitored the quantity and

location of coke formation during the reaction (Methods). As shown in Fig. 4, the MFI nanosheets exhibited not only much slower coke deposition than the conventional MFI (45 versus 170 mg g^{-1} zeolite at 5 days), but also coke formation almost exclusively at the external surface (that is, mesopores) while the conventional MFI zeolite showed major coke formation inside the micropores. Coke deposition within micropores causes more effective catalyst deactivation than external coke formation^{27,28} because internally deposited coke can cover the catalytically active acid sites and also block micropores already at low coking levels; in contrast, external coke causes relatively little hindrance to diffusion unless it covers the entire external catalyst surface. We therefore propose that the long catalytic lifetime of the MFI nanosheets is due to the slow deposition of coke exclusively at external zeolite surfaces, which arises because of facile mass transfer of coke precursors out of the zeolite micropores. Although we observed slow catalyst deactivation in the methanol-to-gasoline case study, it is expected that the MFI nanosheets would generally show high catalyst lifetime in various reactions²⁹.

The MFI nanosheets exhibited excellent thermal stability (Supplementary Fig. 2), hydrothermal stability (Supplementary Fig. 3) and strong acidity (Supplementary Fig. 4 and Supplementary Table 1), which are important for many catalytic applications.²⁷ Al magic-angle spinning NMR spectra indicated that approximately 50% of the initial tetrahedral Al was retained in the zeolite framework even after being heated in 100% steam at 700°C (Supplementary Fig. 5). In initial experiments, we also used the present synthesis strategy to create nanosheets of the zeolite MTW (Supplementary Figs 6 and 7). This suggests that the structure-directing strategy that targets the mesoporous and microporous length scales simultaneously is fairly general, and that it can be extended to other zeolite structures and zeotype materials through the design of suitable bifunctional surfactants. The next challenge is to synthesize such porous materials in the form of continuous films or membranes for advanced applications in catalysis, adsorption, separation and sensor technologies.

Table 1 | Catalytic conversion of bulky molecules over MFI zeolites

Reactions	Conventional MFI (Si/Al = 41)	Multilamellar MFI nanosheets (Si/Al = 48)	Unilamellar MFI nanosheets (Si/Al = 53)
Cracking of branched polyethylene (HDPE)	27	45	85
 Flavanone Chalcone	16 (50/50/0)*	48 (62/28/10)*	76 (64/31/5)*
 Diacetal	42	86	86

Catalytic activities were compared on the basis of the same weight of catalyst (see Methods for reaction conditions). *The numbers in parentheses indicate percentage selectivity: (flavanone/chalcone/others). All other numbers indicate the percentage reactant conversion, reproducible within 3% over three runs. HDPE, high-density polyethylene.

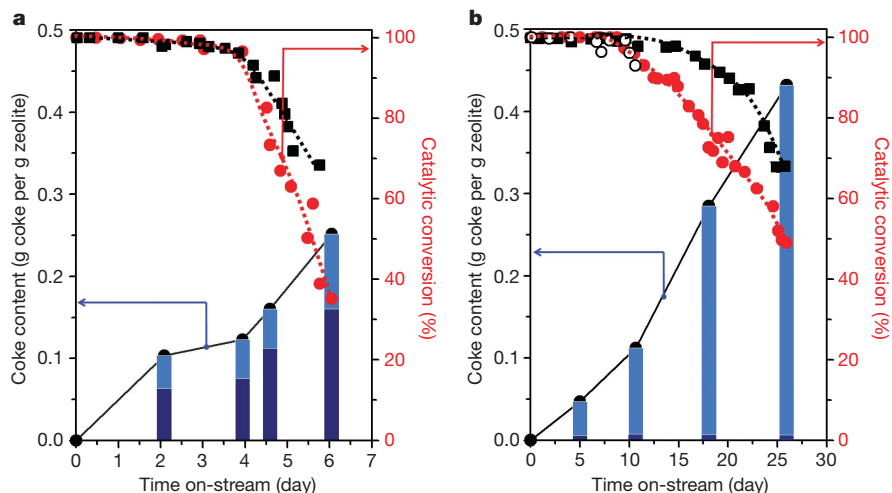


Figure 4 | Coke deposition in MFI zeolite catalysts during methanol-to-gasoline conversion. **a**, Conventional MFI zeolite. **b**, Unilamellar MFI zeolite. The unilamellar MFI zeolite exhibits a dramatically increased catalytic lifetime compared with its conventional counterpart, which is related to the preferential formation of coke in mesopores. Catalytic conversion over the unilamellar MFI was repeatedly investigated using three different synthesis batches (red circles, black squares, open circles, respectively). The catalytic measurement for conventional zeolite was repeated twice using the same sample (red circles and black squares). The solid black lines and the dotted red and black lines are guides to the eye. Dark blue bars indicate internal (inside the micropores of the zeolite) coke content, and light blue bars indicate external coke content.

METHODS SUMMARY

The C_{22-6-6} surfactant was synthesized in the bromide form, that is, $C_{22-6-6}Br_2$. This surfactant was mixed with tetraethylorthosilicate (or sodium silicates), aluminium sulphate, NaOH, H_2SO_4 and distilled water, to give a molar composition of $30 Na_2O:1 Al_2O_3:100 SiO_2:10 C_{22-6-6}Br_2:18 H_2SO_4:4,000 H_2O$. This mixture was heated at $150^\circ C$ for 5 days in an autoclave (set on 'tumbling'), to obtain the multilamellar MFI zeolite. The unilamellar MFI zeolite was synthesized at $1 Al_2O_3:100 SiO_2:15 C_{22-6-6}(OH)_2:3 H_2SO_4:6,000 H_2O$, at $150^\circ C$ for 11 days, under sodium-free conditions. The hydroxide form of the surfactant was prepared through the anion exchange treatment of $C_{22-6-6}Br_2$. The conventional MFI zeolite (ZSM-5) used in this work was purchased from Zeolyst. All the zeolite samples possessed similar Si/Al ratios of 41–53. All catalytic reactions were carried out after converting zeolites into the H^+ form.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 February; accepted 13 July 2009.

- Cundy, C. S. & Cox, P. A. The hydrothermal synthesis of zeolites: history and development from the earliest days to the present time. *Chem. Rev.* **103**, 663–701 (2003).
- Corma, A. From microporous to mesoporous molecular sieve materials and their use in catalysis. *Chem. Rev.* **97**, 2373–2419 (1997).
- Corma, A. State of the art and future challenges of zeolites as catalysts. *J. Catal.* **216**, 298–312 (2003).
- Egeblad, K., Christensen, C. H., Kustova, M. & Christensen, C. H. Templating mesoporous zeolites. *Chem. Mater.* **20**, 946–960 (2008).
- Tao, Y., Kanoh, H., Abrams, L. & Kaneko, K. Mesopore-modified zeolites: preparation, characterization, and applications. *Chem. Rev.* **106**, 896–910 (2006).
- Tosheva, L. & Valtchev, V. P. Nanozeolites: synthesis, crystallization mechanism, and applications. *Chem. Mater.* **17**, 2494–2513 (2005).
- Corma, A., Fornes, V., Pergher, S. B., Maesen, ThLM & Buglass, J. G. Delaminated zeolite precursors as selective acidic catalysts. *Nature* **396**, 353–356 (1998).
- Corma, A., Fornés, V., Martínez-Triguero, J. & Pergher, S. B. Delaminated zeolites: combining the benefits of zeolites and mesoporous materials for catalytic uses. *J. Catal.* **186**, 57–63 (1999).
- Corma, A., Diaz, U., Domine, M. E. & Fornés, V. New aluminosilicate and titanosilicate delaminated materials active for acid catalysis, and oxidation reactions using H_2O_2 . *J. Am. Chem. Soc.* **122**, 2804–2809 (2000).
- Holland, B. T., Abrams, L. & Stein, A. Dual templating of macroporous silicates with zeolitic microporous frameworks. *J. Am. Chem. Soc.* **121**, 4308–4309 (1999).
- Jacobsen, C. J. H., Madsen, C., Houzvicka, J., Schmidt, I. & Carlsson, A. Mesoporous zeolite single crystals. *J. Am. Chem. Soc.* **122**, 7116–7117 (2000).
- Tao, Y. S., Kanoh, H. & Kaneko, K. ZSM-5 monolith of uniform mesoporous channels. *J. Am. Chem. Soc.* **125**, 6044–6045 (2003).
- Fan, W. *et al.* Hierarchical nanofabrication of microporous crystals with ordered mesoporosity. *Nature Mater.* **7**, 984–991 (2008).
- Choi, M. *et al.* Amphiphilic organosilane-directed synthesis of crystalline zeolite with tunable mesoporosity. *Nature Mater.* **5**, 718–723 (2006).
- Wang, H. & Pinnavaia, T. J. MFI zeolite with small and uniform intracrystal mesopores. *Angew. Chem. Int. Edn Engl.* **45**, 7603–7606 (2006).
- Serrano, D. P., Aguado, J., Escala, J. M., Rodríguez, J. M. & Peral, Á. Hierarchical zeolites with enhanced textural and catalytic properties synthesized from organofunctionalized seeds. *Chem. Mater.* **18**, 2462–2464 (2006).
- Xiao, F. S. *et al.* Catalytic properties of hierarchical mesoporous zeolites templated with a mixture of small organic ammonium salts and mesoscale cationic polymers. *Angew. Chem. Int. Edn Engl.* **45**, 3090–3093 (2006).

- van Donk, S., Janssen, A. H., Bitter, J. H. & de Jong, K. P. Generation, characterization, and impact of mesopores in zeolite catalysts. *Catal. Rev.* **45**, 297–319 (2003).
- Groen, J. C. *et al.* Creation of hollow zeolite architectures by controlled desilication of Al-zoned ZSM-5 crystals. *J. Am. Chem. Soc.* **127**, 10792–10793 (2005).
- Groen, J. C. *et al.* Direct demonstration of enhanced diffusion in mesoporous ZSM-5 zeolite obtained via controlled desilication. *J. Am. Chem. Soc.* **129**, 355–360 (2007).
- Gao, Y., Yoshitake, H., Wu, P. & Tatsumi, T. Controlled detitanation of ETS-10 materials through the post-synthetic treatment and their applications to the liquid-phase epoxidation of alkenes. *Micropor. Mesopor. Mater.* **70**, 93–101 (2004).
- Pavel, C. C. & Schmidt, W. Generation of hierarchical pore systems in the titanosilicate ETS-10 by hydrogen peroxide treatment under microwave irradiation. *Chem. Commun.* **8**, 882–884 (2006).
- Ryoo, R., Choi, M. & Na, K. Regularly stacked multilamellar and randomly arranged unilamellar zeolite nanosheets, and their analogue materials whose framework thickness were corresponding to one unit cell size or less than 10 unit cell size. Korean patent (applied, 2009).
- Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartuli, J. C. & Beck, J. S. Ordered mesoporous molecular-sieves synthesized by a liquid-crystal template mechanism. *Nature* **359**, 710–712 (1992).
- Beck, J. S. *et al.* Molecular or supramolecular templating: defining the role of surfactant chemistry in the formation of microporous and mesoporous molecular sieves. *Chem. Mater.* **6**, 1816–1821 (1994).
- Shetti, V. N., Kim, J., Srivastava, R., Choi, M. & Ryoo, R. Assessment of the mesopore wall catalytic activities of MFI zeolite with mesoporous/microporous hierarchical structures. *J. Catal.* **254**, 296–303 (2008).
- Bibby, D. M., Milestone, N. B., Patterson, J. E. & Aldridge, L. P. Coke formation in zeolite ZSM-5. *J. Catal.* **97**, 493–502 (1986).
- Bibby, D. M., Howe, R. F. & McLellan, G. D. Coke formation in high-silica zeolites. *Appl. Catal. Gen.* **93**, 1–34 (1992).
- Srivastava, R., Choi, M. & Ryoo, R. Mesoporous materials with zeolite framework: remarkable effect of the hierarchical structure for retardation of catalyst deactivation. *Chem. Commun.* **43**, 4489–4491 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was mainly supported by the National Honor Scientist Program of the Ministry of Education, Science and Technology in Korea. Electron microscopic studies were performed with help from the measurement and analysis team at the National NanoFab Center, from the research supporting team at KAIST, and also at the Electron Microscopy Center (EMC) at Stockholm University with support from the Knut and Alice Wallenberg Foundation. Y.S. and O.T. thank the Swedish Research Council (VR) and the Japan Science and Technology Agency (JST), respectively.

Author Contributions R.R. initiated and managed the whole project, including the intuitive design of the structure-directing agents. M.C. and K.N. did synthesis and characterization (SEM, TEM, X-ray diffraction and adsorption). K.N. synthesized the multilamellar zeolite, and M.C. synthesized the unilamellar zeolite later. J.K. performed the catalytic testing. Y.S. and O.T. did high-resolution SEM and TEM investigations.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.R. (rryoo@kaist.ac.kr).

METHODS

Synthesis of organic surfactant. $C_{22-6-6}Br_2$ was synthesized as follows: 39.0 g (0.100 mol) 1-bromodocosane (TCI) and 172 g (1.000 mol) N,N,N',N' -tetramethyl-1,6-diaminohexane (Aldrich) were dissolved in 1,000 ml acetonitrile/toluene mixture (1:1 vol/vol) and heated at 70 °C for 10 h. After cooling to room temperature, the product was filtered, washed with diethyl ether, and dried in a vacuum oven at 50 °C. 56.2 g (0.100 mol) of the product and 24.6 g (0.200 mol) 1-bromohexane (Aldrich) were dissolved in 300 ml acetonitrile and refluxed for 10 h. After cooling to room temperature, the product was filtered, washed with diethyl ether, and dried in a vacuum oven at 50 °C.

Synthesis of nanosheet MFI zeolite. In a typical synthesis of multilamellar MFI nanosheets, tetraethylorthosilicate (TEOS, from TCI), $Al_2(SO_4)_3 \cdot 18H_2O$ (Aldrich), NaOH, $C_{22-6-6}Br_2$, H_2SO_4 and distilled water were mixed to obtain a gel composition of 30 Na_2O :1 Al_2O_3 :100 SiO_2 :10 $C_{22-6-6}Br_2$:18 H_2SO_4 :4,000 H_2O . Water glass (an aqueous solution of sodium silicate, $SiO_2/Na = 1.75$, 29 wt% SiO_2) may be used as a silica source instead of TEOS. The resultant gel was transferred to a Teflon-coated stainless-steel autoclave, and heated at 150 °C for 5 days with the autoclave set to tumbling at 60 r.p.m. After crystallization, the zeolite product was filtered, washed with distilled water and dried at 120 °C. The product was calcined at 550 °C for 4 h under flowing air. For synthesis of unilamellar MFI, $C_{22-6-6}Br_2$ was converted to $C_{22-6-6}(OH)_2$ by passing aqueous solution through a column packed with anion exchange resin (MTO-Dowex SBR LCNG OH form, Supelco). The resultant solution contained 13 wt% $C_{22-6-6}(OH)_2$. The $C_{22-6-6}(OH)_2$ solution, TEOS, $Al_2(SO_4)_3 \cdot 18H_2O$ and distilled water were mixed to obtain a gel composition of 1 Al_2O_3 :100 SiO_2 :15 $C_{22-6-6}(OH)_2$:3 H_2SO_4 :6,000 H_2O . The mixture was transferred to a Teflon-coated stainless-steel autoclave, and heated at 150 °C for 11 days with the autoclave set to tumbling at 60 r.p.m.

Characterization. X-ray diffraction patterns were taken with a Rigaku Multiflex diffractometer equipped with $CuK\alpha$ radiation (40 kV, 40 mA). SEM images were taken with a JEOL JSM-7401F at a low landing energy (0.3–0.6 keV, in gentle-beam mode). The samples were mounted without crashing and metal coating. TEM images were obtained with a JEOL JEM-3010 with accelerating voltage of 300 kV ($C_s = 0.6$ mm, point resolution 0.17 nm). N_2 adsorption isotherms were measured at the temperature of liquid nitrogen with an ASAP2020 volumetric adsorption analyser. The Brunauer–Emmett–Teller equation was used to

calculate the apparent surface area from the adsorption data obtained at P/P_0 between 0.1 and 0.3. P , pressure; P_0 , standard pressure.

Catalytic reactions. For catalytic reactions, all MFI zeolites synthesized in the present work were NH_4^+ -ion exchanged with a 1 M NH_4NO_3 solution three times in all (NH_4NO_3 /zeolite Al = 10, each time). The zeolite samples were converted to the H^+ form through calcination in air at 550 °C. An MFI zeolite sample in NH_4^+ form was purchased from Zeolyst (sample codes CBV 8014, Si/Al = 41). The zeolite was also calcined at 550 °C. This zeolite is referred to as conventional zeolite.

The catalytic reactions involving large molecules were carried out and analysed, following methods reported in the literature²⁶. Cracking of branched polyethylene was performed in a Pyrex batch reactor equipped with an overhead stirrer. 10 g of polyethylene were placed in the reactor and melted at 350 °C. After the addition of 0.1 g of catalyst, the reactor temperature was further increased to 380 °C. During the reaction, N_2 gas was passed through the reactor at a rate of 40 ml min^{-1} . After 30 min reaction, the reaction yield was calculated from the mass change. Protection of benzaldehyde with pentaerythritol was carried out using a Pyrex batch reactor (EYELA chemistation) equipped with a reflux condenser. 1.06 g benzaldehyde (10 mmol), 0.68 g pentaerythritol (5 mmol), 4 ml toluene and 20 mg catalyst were placed into the Pyrex reactor and heated under stirring for 4 h at 120 °C. Condensation of 2-hydroxyacetophenone with benzaldehyde was carried out by heating a mixture containing benzaldehyde (0.75 g, 7 mmol), 2-hydroxyacetophenone (0.48 g, 3.5 mmol) and 50 mg catalyst at 150 °C for 14 h.

The methanol-to-gasoline reaction was performed at 400 °C in a fixed-bed Pyrex reactor (inner diameter, 13 mm) using 100 mg of catalyst. Before reaction, catalysts were activated at 550 °C for 2 h in a flowing air (30 ml min^{-1}). Methanol (99.6%) vapour was introduced by passing N_2 flow (50 ml min^{-1}) through a saturation evaporator at 30 °C (weight hourly space velocity = 11.0 h^{-1}). By considering oxygenates (methanol and dimethylether) as unconverted species, conversion was calculated by gas chromatography analysis. After prolonged reaction time, the used catalyst was collected and coke content was analysed by thermogravimetric analysis (TA Instrument). The coke formation inside micropores (internal coke) was calculated from a decrease in micropore volume (determined by N_2 adsorption), assuming coke density to be 1.22 $g\ cm^{-3}$ (ref. 27). The coke content deposited on the external surface was calculated by subtracting the internal coke content from the total coke content.

LETTERS

Fluctuations in Precambrian atmospheric oxygenation recorded by chromium isotopes

Robert Frei¹, Claudio Gaucher^{1,2}, Simon W. Poulton³ & Don E. Canfield⁴

Geochemical data^{1–4} suggest that oxygenation of the Earth's atmosphere occurred in two broad steps. The first rise in atmospheric oxygen is thought to have occurred between ~2.45 and 2.2 Gyr ago^{1,5}, leading to a significant increase in atmospheric oxygen concentrations and concomitant oxygenation of the shallow surface ocean. The second increase in atmospheric oxygen appears to have taken place in distinct stages during the late Neoproterozoic era (~800–542 Myr ago)^{3,4}, ultimately leading to oxygenation of the deep ocean ~580 Myr ago³, but details of the evolution of atmospheric oxygenation remain uncertain. Here we use chromium (Cr) stable isotopes from banded iron formations (BIFs) to track the presence of Cr(VI) in Precambrian oceans, providing a time-resolved picture of the oxygenation history of the Earth's atmosphere–hydrosphere system. The geochemical behaviour of Cr is highly sensitive to the redox state of the surface environment because oxidative weathering processes produce the oxidized hexavalent [Cr(VI)] form. Oxidation of reduced trivalent [Cr(III)] chromium on land is accompanied by an isotopic fractionation, leading to enrichment of the mobile hexavalent form in the heavier isotope. Our fractionated Cr isotope data indicate the accumulation of Cr(VI) in ocean surface waters ~2.8 to 2.6 Gyr ago and a likely transient elevation in atmospheric and surface ocean oxygenation before the first great rise of oxygen 2.45–2.2 Gyr ago (the Great Oxidation Event)^{1,5}. In ~1.88-Gyr-old BIFs we find that Cr isotopes are not fractionated, indicating a decline in atmospheric oxygen. Our findings suggest that the Great Oxidation Event did not lead to a unidirectional stepwise increase in atmospheric oxygen. In the late Neoproterozoic, we observe strong positive fractionations in Cr isotopes ($\delta^{53}\text{Cr}$ up to +4.9‰), providing independent support for increased surface oxygenation at that time, which may have stimulated rapid evolution of macroscopic multicellular life^{3,4,6}.

The mobile Cr(VI) anion (HCrO_4^-) is the most thermodynamically stable Cr form in equilibrium with present-day air. Oxidation of Cr(III) to Cr(VI) in soils depends upon the co-occurrence of Cr(III) (bound most commonly as FeCr_2O_4) and manganese oxides (catalysing Cr(III) oxidation). Once mobilized during oxidative weathering, Cr(VI) is mobile as either chromate (CrO_4^{2-} ; alkaline pH) or bichromate (HCrO_4^- ; acidic pH) ions, entering the oceans via riverine transport⁷. There is a considerably smaller input of Cr from atmospheric and hydrothermal vent sources. In today's oceans, total dissolved Cr concentrations are in the range of 2 to 10 nM with a relatively short residence time of ~2.5 to 4×10^4 years⁸.

Cr(VI) can be reduced to Cr(III) by microbes⁹ and by aqueous Fe(II) or Fe(II)-bearing minerals¹⁰ (see equation (1)). Indeed, the oxidation of Fe(II) (aq) by Cr(VI) proceeds faster than with oxygen, even under well-aerated, high-pH conditions¹¹. This means that in the presence of Fe(II), Cr(IV) is efficiently reduced to Cr(III). The

Cr(III) is subsequently and effectively scavenged into Fe(III)–Cr(III) oxyhydroxides¹² owing to the very low solubility of $\text{Fe}_2(\text{OH})_2\text{Cr}_2\text{O}_7$ solids¹³. Some Cr(III) can be regenerated and lost from sediments as a result of Fe oxide reduction, but, as on land, the Cr(III) is reoxidized rapidly¹⁴ to Cr(VI) in a catalytic reaction with MnO_2 (ref. 7):



At equilibrium, the Cr(VI)O_4^{2-} anion is enriched by up to 7‰ at room temperature in ^{53}Cr compared to coexisting compounds containing Cr(III) (we use the delta notation relative to the certified National Bureau of Standards Cr reference standard SRM 979, defined as $\delta^{53}\text{Cr} = 1,000 \times [({}^{53}\text{Cr}/{}^{52}\text{Cr})_{\text{sample}}/({}^{53}\text{Cr}/{}^{52}\text{Cr})_{\text{SRM 979}} - 1]$ (see ref. 15). Therefore, subsurface aqueous environments will have positive $\delta^{53}\text{Cr}$ values¹⁶. Although the isotopic composition of Cr in sea water has not yet been measured, the positive groundwater Cr(VI) signal should be transferred to the sea, because subsequent adsorption of Cr onto particles (as might occur in soils and rivers) produces no isotope effect¹⁷. The microbial reduction of Cr(VI) generates isotopic shifts of up to –4.1‰, comparable to those produced during abiotic reduction^{9,10}. This will potentially enrich the heavier isotope in the remaining, unreacted, dissolved Cr(VI). However, because of the efficient sequestration of Cr(VI) during Cr reduction and subsequent precipitation of Cr(III) with Fe-oxyhydroxides, the stable Cr isotope signatures of chemically precipitated Fe(III)-rich sediments should mirror the sea water from which the Fe oxides precipitated. The surface chemistry of Cr and its stable isotope geochemistry are summarized in Fig. 1.

The prerequisite for Cr isotopes to record the presence of Cr(VI) in sea water is a predominance of dissolved Fe(II), which acts as the reductant. Therefore, the isotopic composition of Cr in ancient iron-rich sediments should provide a first-order proxy for the presence of Cr(VI) in ancient surface waters, and thus the history of the oxidative weathering of Cr on land. This approach should be relatively insensitive to the type of iron-rich chemical sediments and the palaeoenvironment in which these were deposited.

Oxidation and solubilization of Cr from soils is strongly dependent on the presence of MnO_2 , which is stable under elevated oxygen fugacities, so the pathway of Cr to the oceans in the early Precambrian would have been limited by the absence of Mn(IV) under low atmospheric oxygen pressures. The geochemical behaviour of Cr in sea water is therefore highly sensitive to levels of atmospheric oxygen (see Supplementary Information for further details).

We analysed $\delta^{53}\text{Cr}$ values of numerous Precambrian BIFs (Supplementary Table 1), from which we delineate six stages of Cr cycling (Fig. 2). During stage 1, comprising BIFs deposited during the Archaean from 3.7–2.8 Gyr, the $\delta^{53}\text{Cr}$ values are unfractionated

¹Institute of Geography and Geology and Nordic Center for Earth Evolution (NordCEE), University of Copenhagen, Øster Voldgade 10, 1350 Copenhagen, Denmark. ²Departamento de Geología, Facultad de Ciencias, Iguá 4225, 11400 Montevideo, Uruguay. ³School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK. ⁴Nordic Center for Earth Evolution (NordCEE) and Institute of Biology, University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark.

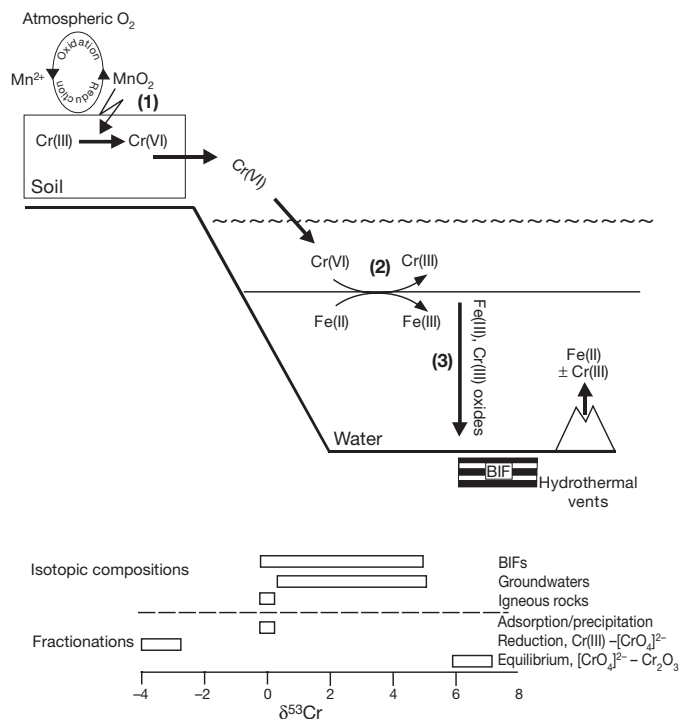


Figure 1 | Schematic of the surface chemistry of chromium. (1): Oxidation of Cr(III) in soils is catalysed by MnO_2 and positively fractionated Cr(VI)^{10,16} enters the aquatic phase (groundwater, rivers) mainly as HCrO_4^- complexes, and eventually enters the ocean. (2): Abiotic reduction of Cr(VI) by upwelling Fe(II)¹⁰ is efficient, fast and complete. (3): Subsequent scavenging of Cr by Fe–Cr oxyhydroxides is a major removal pathway of Cr into the sedimentary environment. The positively fractionated Cr ($\delta^{53}\text{Cr} \approx -0.3\text{‰}$ to 4.9‰ ; this study) in BIFs and Fe-rich cherts thereby mirrors the riverine Cr(VI) input. Biotic (bacterial) reduction of Cr(VI) has recently been reported with Cr isotopic shifts ($\Delta^{53}\text{Cr}_{\text{Cr(III)}-\text{Cr(VI)}}$) of up to -4.1‰ (ref. 9). Adsorption and complexation of Cr(III), and to a lesser extent Cr(VI), on or with organic and inorganic particles is not accompanied by Cr isotopic shifts¹⁷. Cr(III) input into sea water by hydrothermal vents is considered small and the Cr isotopic composition of this fraction possibly reflects the $\delta^{53}\text{Cr}$ values of $\sim 0.15\text{‰}$ typical of magmatic high-temperature reservoirs¹⁸. Back-transformation of Cr(III) to Cr(VI) from sediments to sea water is again only possible when catalysed by MnO_2 .

compared to the Earth's high-temperature igneous reservoirs¹⁸. This indicates a lack of oxidative continental weathering during most of the Archaean. In stage 2, during the latest Archaean and before the Great Oxidation Event (GOE) in the early Proterozoic (~ 2.8 – 2.45 Gyr ago), four out of seven BIFs show positively fractionated $\delta^{53}\text{Cr}$ values of $+0.04$ to $+0.29\text{‰}$ (Fig. 2, Supplementary Table 1). We interpret these enrichments to reflect transient occurrences of slightly elevated oxygen. This is consistent with Mo concentration and S isotope evidence for ephemeral surface water (and possibly atmospheric) oxygenation in the run-up to the GOE^{1,5,19–22}, although our results would suggest that Cr(VI) was mobilized up to 300 Myr before the GOE.

Stage 3, from ~ 2.45 to 1.9 Gyr, is defined by rare BIF deposition, particularly during the GOE itself. This apparent absence of BIFs could reflect a transition from Fe-rich to Fe-poor oceans. It is interesting that the BIFs just predating the GOE (the youngest samples in stage 2 deposited about 2.5 – 2.4 Gyr ago) show little Cr enrichment, and were presumably deposited just before the GOE and the associated rise in atmospheric O_2 . Major BIFs were again deposited in North America, India and Australia²³ at around 2.1 Gyr (our samples come from South Dakota and the Bastar Craton, Supplementary Table 1). These stage 3 BIFs display some positive $\delta^{53}\text{Cr}$ values when compared to the range of magmatic values and compared to stage 1 BIFs (Fig. 2), and roughly coincide with BIFs characterized by positively fractionated $\delta^{56}\text{Fe}$ values² of sedimentary pyrite, interpreted to reflect an increase in the precipitation of iron sulphides relative to iron oxides in a redox stratified ocean. These fractionated Cr isotopes follow the GOE by some 200 to 300 Myr and would be consistent with concomitant elevated levels of atmospheric oxygenation. However, measured Cr fractionations are lower than the elevated fractionations observed in stage 2, a time where consensus would argue for very low levels of atmospheric O_2 punctuated by occasional 'whiffs' of oxygen^{19–21}. As we shall see below, there is evidence that atmospheric oxygen declined after the GOE. Therefore, lower fractionations in late stage 3 could, in part, have resulted from a return to reduced oxygen levels.

Indeed, the oldest of our stage 4 samples, which come from the Gunflint Iron Formation (Ontario, Canada), do not show any positively fractionated $\delta^{53}\text{Cr}$ values (Fig. 2; Supplementary Table 1). This is further evidence for a decrease in atmospheric oxygen after the GOE. Supporting evidence for low atmospheric oxygen comes from the precipitation of iron oxides in the high-energy near-shore region as observed in the 1.88 -Gyr-old Gunflint Iron Formation²⁴. This required the transport of dissolved Fe^{2+} over a broad continental shelf to the palaeoshoreline and atmospheric O_2 concentrations no greater

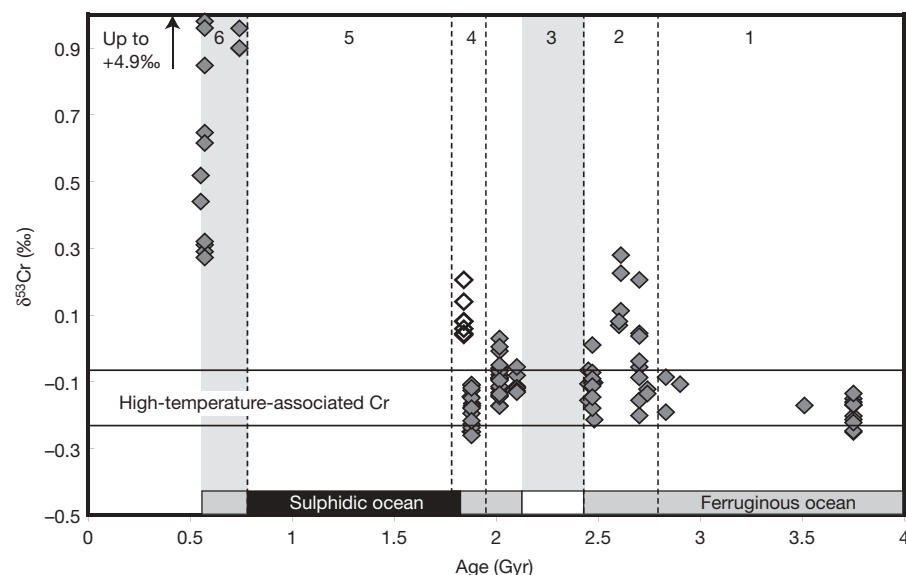


Figure 2 | Graph showing the key aspects of the Precambrian history of hexavalent chromium in sea water. $\delta^{53}\text{Cr}$ values (grey filled diamond symbols) for BIF versus age (22 localities in total; high values up to $+4.9\text{‰}$ from Neoproterozoic Fe-rich cherts plot outside the graph; data in Supplementary Table 1). Six stages (separated by dashed vertical lines) are identified and compared to the ocean deep water chemistry²⁷. The light-grey shaded fields depict the first and second GOE, respectively, as defined by other redox-sensitive tracers. Open diamonds designate data from the upper Gunflint Iron Formation (Ontario, Canada) which is transitional into the overlying Rove Formation (detail in Fig. 3). The horizontal rectangular field outlines the $\delta^{53}\text{Cr}$ values of magmatic Cr(III)-rich ores and minerals formed under high temperatures¹⁸. Data are reported using the delta notation relative to the certified National Bureau of Standards Cr reference standard SRM 979 (see Methods). Error bars associated with the individual symbols for $\delta^{53}\text{Cr}$ values are $<0.1\text{‰}$ (2σ level; full analytical data are available in Supplementary Table 1).

than about 0.1% of the present levels²⁵. Furthermore, we speculate that reduced oxygen levels may have inhibited the weathering flux of sulphate to the oceans, encouraging the return of marine ferruginous conditions and BIF deposition. Does this mean that oxygen concentrations were reduced to levels comparable to the early Archaean? We see no evidence to suggest this. Mass independent sulphur isotope fractionations persist through the Archaean²⁶, but are not observed in stage 4 sediments²⁵. We suggest that chromium systematics are very sensitive to oxygen, but not linearly so, such that factors controlling MnO₂ availability (as it relates to Cr(III) oxidation) might also be important. More work on these systematics will settle this issue.

Higher ⁵³Cr isotope values are observed during the very final stages of Gunflint deposition (open diamonds), implying a subsequent increase in oxygen concentrations (Fig. 3; Supplementary Table 1). This represents the time immediately before the likely development of widespread sulphidic oceanic conditions, which are thought to have

persisted throughout much of the Mesoproterozoic^{27–29}. The traditional explanation for the development of sulphidic conditions calls for an increase in the flux of sulphate to the oceans (and hence increased rates of sulphide production by bacterial sulphate reduction) due to enhanced oxidative weathering of continental sulphide minerals as a result of the GOE²⁷. It has never been clear, however, why it would take several hundred million years for the sulphide flux to overwhelm the hydrothermal Fe(II) flux, thus allowing sulphidic conditions eventually to develop at ~1.84 Gyr (ref. 29). Our Cr isotope data implies that atmospheric O₂ concentrations fluctuated over this period, and the onset of sulphidic conditions at ~1.84 Gyr is a consequence of an increased sulphate flux arising owing to a previously unrecognized rise in atmospheric O₂ during the final stages of Palaeoproterozoic BIF deposition.

Stage 5, between ~1.8 and 0.75 Gyr, comprises the Mesoproterozoic period in which sulphidic oceans predominated²⁷ and during which BIF deposition was largely prevented because of the preferential titration of Fe²⁺ by HS[–].

Stage 6 comprises the late Neoproterozoic era between ~750 Myr and the Precambrian–Cambrian boundary at 542 Myr. BIFs deposited during this stage include the 755–730-Myr-old Rapitan BIF, deposited in a glaciomarine setting during the early Cryogenian ('Sturtian') glaciation, and BIF- and Fe-bearing cherts of the ~570–550-Myr-old Yerbal and Cerro Espuelitas formations (Arroyo del Soldado Group, Uruguay) which were deposited after the Gaskiers glaciation³⁰. All of these BIFs record strongly positively fractionated Cr isotopes, with $\delta^{53}\text{Cr}$ values ranging from 0.9‰ to 4.9‰. These high values provide independent support for Late Neoproterozoic oxygenation, which further points to a causal link between this oxygenation and the emergence of the Ediacara biota³ and bilateral motile animals.

Our Cr isotope record provides new and complementary insights into the history of Precambrian biospheric oxygenation. The data highlights fine-scale fluctuations in the oxygenation of the ocean and atmosphere through time, and we foresee that combining Cr isotope systematics with information from other redox-sensitive elements, such as C, S, Fe and Mo, will greatly enhance our understanding of the complex history of chemical and biological evolution on the early Earth.

METHODS SUMMARY

Individual mesobands of BIF samples were isolated from one-centimetre-thick slices of hand specimens or drill core pieces and subsequently milled in an agate mortar. Rock powder aliquots (amounts adjusted to yield 2–5 µg Cr in the final separate) were spiked with an adequate amount of a ⁵⁰Cr–⁵⁴Cr double spike and digested in HF:HNO₃ mixtures in closed teflon vials on a hot plate at 150 °C. The samples were then taken up in 6 M hydrochloric acid and passed through an exchange column charged with 6 ml Dowex AG 1 × 12 anion resin to remove Fe. Oxidation of Cr(III) to Cr(VI) in dilute hydrochloric acid was then achieved by addition of (NH₄)₂S₂O₈ as an oxidizing agent on a hot plate at 130 °C. In a second chromatographic separation, the dilute Cr(VI) solutions were processed over chromatographic columns charged with Dowex AG 1 × 8 anion resin. Release of Cr from the anion resin was achieved by reduction to Cr(III) with the help of 2 M nitric acid and hydrogen peroxide. All Cr isotope measurements were performed on a IsotopX/GV IsoProbe T thermal ionization mass spectrometer equipped with eight Faraday collectors that allow simultaneous collection of all four chromium beams (⁵⁰Cr⁺, ⁵²Cr⁺, ⁵³Cr⁺, ⁵⁴Cr⁺) together with ⁴⁹Ti⁺, ⁵¹V⁺ and ⁵⁶Fe⁺ as monitors for small interferences of these masses on ⁵⁰Cr and ⁵⁴Cr. Cr separates were measured from Re-filaments at 1,000–1,100 °C and loaded with ultraclean water into a mixture of 3 µl silica gel, 0.5 µl 0.5 M H₃BO₃ and 0.5 µl of 0.5 M H₃PO₄. Every separate was analysed up to six times with minimum ⁵²Cr beam intensities of 400 mV. Data are reported relative to the certified Cr isotope standard NIST SRM 979 (see online-only Methods). Within-run precisions of the sample $\delta^{53}\text{Cr}$ values were ±0.08‰ (2σ) or better.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 February; accepted 30 June 2009.

1. Bekker, A. *et al.* Dating the rise of atmospheric oxygen. *Nature* **427**, 117–120 (2004).
2. Rouxel, O. J., Bekker, A. & Edwards, K. J. Iron isotope constraints on the Archean and Paleoproterozoic ocean redox state. *Science* **307**, 1088–1091 (2005).

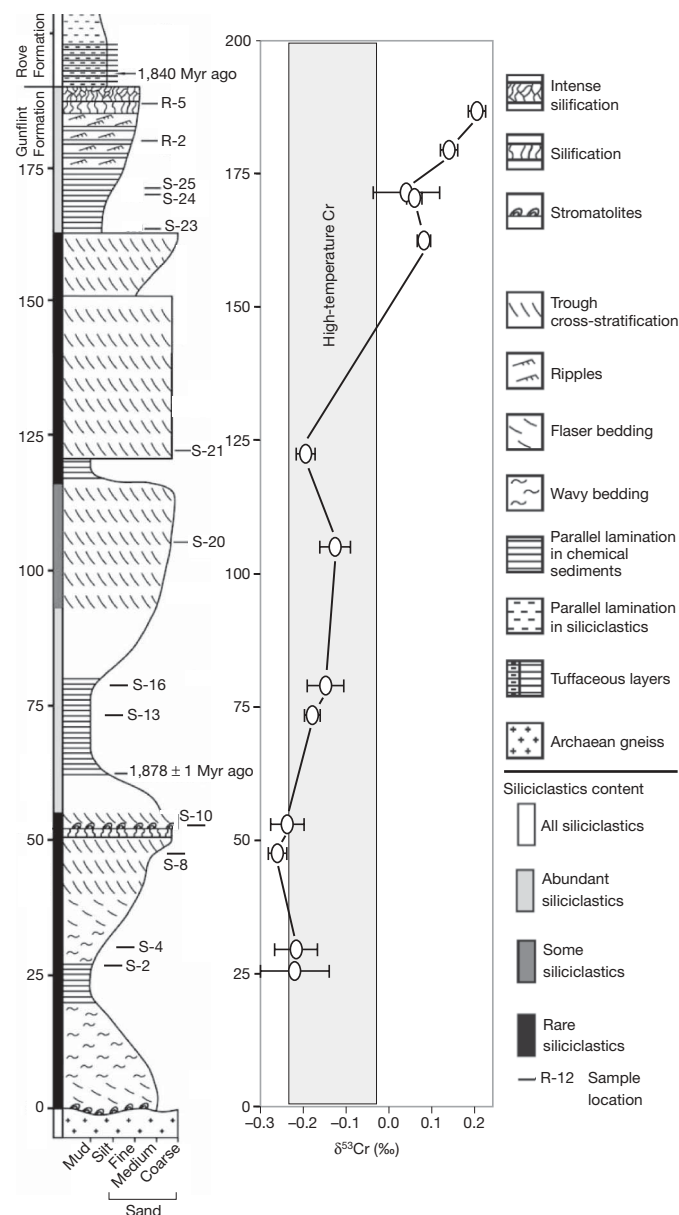


Figure 3 | Stratigraphy of the Gunflint Formation and its transition into the Rove Formation²⁹ and sample horizons. Data show an increase in $\delta^{53}\text{Cr}$ values in the uppermost Gunflint Iron Formation relative to values typical of minerals and ores associated with magmatic (high-temperature) processes¹⁸ (grey filled rectangle in the data log). Error bars correspond to the $2\sigma_m$ of repeat analyses of respective samples (full analytical data are available in Supplementary Table 1).

3. Canfield, D. E., Poulton, S. W. & Narbonne, G. M. Late-Neoproterozoic deep-ocean oxygenation and the rise of animal life. *Science* **315**, 92–95 (2007).
4. Fike, D. A. *et al.* Oxidation of the Ediacaran Ocean. *Nature* **444**, 744–747 (2006).
5. Bekker, A. & Kaufman, A. Oxidative forcing of global climate change: a biogeochemical record across the oldest Paleoproterozoic ice age in North America. *Earth Planet. Sci. Lett.* **258**, 486–499 (2007).
6. Canfield, D. E. & Teske, A. Late Proterozoic rise in atmospheric oxygen concentration inferred from phylogenetic and sulphur-isotope studies. *Nature* **382**, 127–132 (1996).
7. Oze, C., Bird, D. K. & Fendorf, S. Genesis of hexavalent chromium from natural sources in soil and groundwater. *Proc. Natl Acad. Sci. USA* **104**, 6544–6549 (2007).
8. Campbell, J. A. & Yeats, P. A. Dissolved chromium in the northwest Atlantic Ocean. *Earth Planet. Sci. Lett.* **53**, 427–433 (1981).
9. Sikora, E. R., Johnson, T. M. & Bullen, T. D. Microbial mass-dependent fractionation of chromium isotopes. *Geochim. Cosmochim. Acta* **72**, 3631–3641 (2008).
10. Ellis, A. S., Johnson, T. M. & Bullen, T. D. Chromium isotopes and the fate of hexavalent chromium in the environment. *Science* **295**, 2060–2062 (2002).
11. Eary, L. E. & Rai, D. Kinetics of chromate reduction by ferrous-ions derived from hematite and biotite at 25°C. *Am. J. Sci.* **289**, 180–213 (1989).
12. Fendorf, S. E. Surface-reactions of chromium in soils and waters. *Geoderma* **67**, 55–71 (1995).
13. Sass, B. M. & Rai, D. Solubility of amorphous chromium(III)-iron(III) hydroxide solid-solutions. *Inorg. Chem.* **26**, 2228–2232 (1987).
14. Eary, L. E. & Rai, D. Chromate removal from aqueous wastes by reduction with ferrous ion. *Environ. Sci. Technol.* **22**, 972–977 (1988).
15. Schauble, E., Rossman, G. R. & Taylor, H. P. Theoretical estimates of equilibrium chromium-isotope fractionations. *Chem. Geol.* **205**, 99–114 (2004).
16. Izbicki, J. A. *et al.* Chromium, chromium isotopes and selected elements, western Mojave Desert, USA. *Appl. Geochem.* **23**, 1325–1352 (2008).
17. Ellis, A. S., Johnson, T. M. & Bullen, T. D. Using chromium stable isotope ratios to quantify Cr(VI) reduction: lack of sorption effects. *Environ. Sci. Technol.* **38**, 3604–3607 (2004).
18. Schoenberg, R. *et al.* The stable Cr isotope inventory of solid Earth reservoirs determined by double spike MC-ICP-MS. *Chem. Geol.* **249**, 294–306 (2008).
19. Anbar, A. D. *et al.* A whiff of oxygen before the Great Oxidation Event? *Science* **317**, 1903–1906 (2007).
20. Kaufman, A. *et al.* Late Archean biospheric oxygenation and atmospheric evolution. *Science* **317**, 1900–1903 (2007).
21. Wille, M. *et al.* Evidence for a gradual rise of oxygen between 2.6 and 2.5 Ga from Mo isotopes and Re-PGE signatures in shales. *Geochim. Cosmochim. Acta* **71**, 2417–2435 (2007).
22. Kamber, B. S. & Whitehouse, M. J. Micro-scale sulphur isotope evidence for sulphur cycling in the late Archean shallow ocean. *Geobiology* **5**, 5–17 (2007).
23. Isley, A. E. & Abbott, D. H. Plume-related mafic volcanism and the deposition of banded iron formation. *J. Geophys. Res.* **104**, 15461–15477 (1999).
24. Canfield, D. E. The early history of atmospheric oxygen: homage to Robert A. Garrels. *Annu. Rev. Earth Planet. Sci.* **33**, 1–36 (2005).
25. Johnston, D. T. *et al.* Evolution of the oceanic sulfur cycle at the end of the Paleoproterozoic. *Geochim. Cosmochim. Acta* **70**, 5723–5739 (2006).
26. Farquhar, J. & Wing, B. A. Multiple sulfur isotopes and the evolution of the atmosphere. *Earth Planet. Sci. Lett.* **213**, 1–13 (2003).
27. Canfield, D. E. A new model for Proterozoic ocean chemistry. *Nature* **396**, 450–453 (1998).
28. Scott, C. *et al.* Tracing the stepwise oxygenation of the Proterozoic ocean. *Nature* **452**, 456–459 (2008).
29. Poulton, S. W., Fralick, P. W. & Canfield, D. E. The transition to a sulphidic ocean similar to 1.84 billion years ago. *Nature* **431**, 173–177 (2004).
30. Gaucher, C. *et al.* Acritarchs of Las Ventanas Formation (Ediacaran, Uruguay): implications for the timing of coeval rifting and glacial events in western Gondwana. *Gondwana Res.* **13**, 488–501 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This study was funded by the Danish Agency for Science, Technology and Innovation and by the Danish National Research Foundation (Danmarks Grundforskningsfond). We thank R. Schoenberg for sharing his Cr double spike with us and for contributions during many thematic and analytical discussions. To add to our own collection, critical samples were provided by C. Klein, A. Polat, P. S. Dahl, S. K. Mondal, H.-J. Hansen and E. F. Duke. Help with the clean-laboratory chemical separation procedures by T. Larsen is acknowledged. T. Leeper helped with the mass spectrometry and kept the mass spectrometer in excellent condition.

Author Contributions R.F. and C.G. collected critical Neoproterozoic samples during fieldwork in Uruguay in 2006. D.E.C. and S.W.P. provided the important sediment samples from the Gunflint and Rove formations. Methods development and thermal ionization mass spectrometer analytical work was conducted by R.F. The manuscript was produced by significant contributions by R.F., D.E.C., S.W.P. and C.G. Furthermore, D.E.C. and S.W.P. provided deeper insights into Proterozoic atmospheric and deep ocean oxygenation models and stimulated discussion of the early Earth system evolution.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.F. (robertf@geo.ku.dk).

METHODS

Individual mesobands of BIF samples were isolated from one-centimetre-thick slices of hand specimens or drill core pieces and subsequently milled in an agate mortar.

For trace elemental analyses, the rock powders were attacked in HBr, dissolved with HF and HNO₃ during addition of HBO₃ (ref. 31), and then dried and redissolved in HNO₃. Trace-element concentrations were determined by solution ICP-MS (inductively coupled plasma mass spectrometry) with a Perkin Elmer ELAN 6100 DRC spectrometer at the Geological Survey of Denmark and Greenland (GEUS), using international standards for calibration. For a comparison of GEUS analytical results on some standards with published values, refer to table 1 in ref. 32.

Rock powder aliquots (amounts adjusted to yield 2–5 µg Cr in the final separate) were spiked with an adequate amount of a ⁵⁰Cr–⁵⁴Cr double spike and digested in HF:HNO₃ mixtures in closed PFA vials on a hot plate at 150 °C. After drying down, the residues were taken up in aqua regia and reheated to 170 °C for a couple of hours to destroy fluoride complexes that may have formed during the digestion. After renewed drying down, the sample was then taken up in 6 M hydrochloric acid for the Cr extraction.

We used an anion exchange chromatography technique adapted from previously published methods^{33,34} with few modifications to separate the Cr of natural samples from the other matrix elements. First, because of the high iron contents of our BIF, we passed the solutions through a cation exchange column charged with 6 ml Dowex AG 1 × 12 in 6 M HCl to remove Fe. Sometimes we passed the samples twice over this column to ensure that Fe was removed quantitatively. In a second chromatographic separation over 1 ml stem columns charged with Dowex AG 1 × 8 anion resin, we cleaned the Cr fractions collected from the Fe-cleanup columns in dilute 0.2 M HCl from other rock matrices. This separation method is based on exchange of chloride ions on the Dowex AG 1 × 8 resin by the Cr(VI) oxyanions¹⁸. After sample digestion and the first cationic exchange Cr is present in its trivalent (Cr(III)) form, so oxidation of Cr(III) to Cr(VI) was achieved using (NH₄)₂S₂O₈ as oxidizing agent³³ on a hot plate at 130 °C. Release of Cr from the anion resin was achieved by reduction to Cr(III) using 2 M HNO₃ and H₂O₂. The procedure yields for Cr in the above described separation method varied between 80–90%, and Cr procedure blanks were in the order of 5–10 ng, which is negligible compared to the amount of Cr separated from the samples studied herein.

The addition of a ⁵⁰Cr–⁵⁴Cr double spike of known isotope composition to a sample before chemical purification allowed accurate correction of both the chemical and the instrumental shifts in Cr isotope abundances^{10,18}. With this method we achieve a 2σ external reproducibility of the δ⁵³Cr value with 1.5 µg Cr loads of the NIST SRM 3112a standard on our IsotopX/GV IsoProbe T thermal ionization mass spectrometer of ±0.05‰ with ⁵²Cr signal intensities of 1 V and of ±0.08‰ for ⁵²Cr beam intensities of 500 mV. The double-spike correction returns Cr isotope compositions of samples as the ‰ difference to the isotope composition of the NIST SRM 3112a Cr standard (which was used for the spike calibration¹⁸, so to maintain inter-laboratory comparability of Cr isotope data,

we recalculate our data of natural samples relative to the certified Cr isotope standard NIST SRM 979 as follows:

$$\delta^{53}\text{Cr}_{\text{sample (SRM 979)}} = \left[\left(\frac{{}^{53}\text{Cr}}{{}^{52}\text{Cr}} \right)_{\text{sample}} / \left(\frac{{}^{53}\text{Cr}}{{}^{52}\text{Cr}} \right)_{\text{SRM 979}} - 1 \right] \times 1,000.$$

All Cr isotope measurements were performed on an IsotopX/GV IsoProbe T thermal ionization mass spectrometer equipped with eight Faraday collectors that allow simultaneous collection of all four chromium beams (⁵⁰Cr⁺, ⁵²Cr⁺, ⁵³Cr⁺, ⁵⁴Cr⁺) together with ⁴⁹Ti⁺, ⁵¹V⁺ and ⁵⁶Fe⁺ as monitors for small interferences of these masses on ⁵⁰Cr and ⁵⁴Cr. Cr separates were measured from Re-filaments at 1,000–1,100 °C and loaded with ultraclean water into a mixture of 3 µl silica gel, 0.5 µl 0.5 M H₃BO₃ and 0.5 µl of 0.5 M H₃PO₄. Every separate was analysed 1–6 times with minimum ⁵²Cr beam intensities of 400 mV, allowing within-run precision of the δ⁵³Cr value of ±0.09‰ or better. To achieve this, we ran the sample over 120 cycles (grouped into 24 blocks of five cycles each) in static mode, and integrated over 10 s with 20 s background (baseline) collection at 0.5 AMU on either side of the peaks. This led to an average analysis time of ~1.5 hours. The final δ⁵³Cr value of a sample was then calculated as the average of the repeated analyses. We spiked our samples with an aliquot of the double spike used by Schoenberg *et al.*¹⁸ in their study of silicates and oxides of magmatic and metamorphic rocks, and employed the double-spike correction developed by their group. The external reproducibility of the NIST SRM 3112a standard over a period of one year, using the ⁵⁰Cr/⁵⁴Cr ratio of Shields³⁵ for mass bias correction, is shown in Supplementary Fig. 1. Our average ⁵³Cr/⁵²Cr and ⁵⁰Cr/⁵²Cr ratios for this standard are 0.113452 ± 50 p.p.m. (*n* = 200, 2σ) and 0.0282095 ± 151 p.p.m. (*n* = 200; 2σ), respectively. These values are indistinguishable from those of the SRM 979 isotopic standard, which we reproduce at a ⁵³Cr/⁵²Cr ratio of 0.1134502 ± 78 p.p.m. (*n* = 100; 2σ) and a ⁵⁰Cr/⁵²Cr ratio of 0.0282089 ± 161 p.p.m. (*n* = 100; 2σ). This coincidence means we did not have to correct our δ⁵³Cr_{NIST SRM 3112a} values to maintain inter-laboratory comparability. The average δ⁵³Cr_{NIST SRM 3112a} is −0.019 ± 0.050‰ (*n* = 32, 2σ). The small deviation from the nominal value of 0‰ is most probably due to a small inaccuracy in the calibration of the Cr isotope composition of the double spike¹⁸.

31. Connelly, J. N. *et al.* A method for purifying Lu and Hf for analyses by MC-ICP-MS using TODGA resin. *Chem. Geol.* **233**, 126–136 (2006).
32. Kalsbeek, F. & Frei, R. The Mesoproterozoic Midsommerso dolerites and associated high-silica intrusions, North Greenland: crustal melting, contamination and hydrothermal alteration. *Contrib. Mineral. Petrol.* **152**, 89–110 (2006).
33. Ball, J. W. & Bassett, R. L. Ion exchange separation of chromium from natural water matrix for stable isotope mass spectrometric analysis. *Chem. Geol.* **168**, 1–2 (2000).
34. Frei, R. & Rosing, M. T. Search for traces of the late heavy bombardment on Earth—results from high precision chromium isotopes. *Earth Planet. Sci. Lett.* **236**, 28–40 (2005).
35. Shields, W. R. *et al.* Absolute isotopic abundance ratios and atomic weight of a reference sample of chromium. *J. Res. Natl Bureau Standards* **A70**, 193–197 (1966).

LETTERS

The importance of niches for the maintenance of species diversity

Jonathan M. Levine¹ & Janneke HilleRisLambers²

Ecological communities characteristically contain a wide diversity of species with important functional, economic and aesthetic value. Ecologists have long questioned how this diversity is maintained^{1–3}. Classic theory shows that stable coexistence requires competitors to differ in their niches^{4–6}; this has motivated numerous investigations of ecological differences presumed to maintain diversity^{3,6–8}. That niche differences are key to coexistence, however, has recently been challenged by the neutral theory of biodiversity, which explains coexistence with the equivalence of competitors⁹. The ensuing controversy has motivated calls for a better understanding of the collective importance of niche differences for the diversity observed in ecological communities^{10,11}. Here we integrate theory and experimentation to show that niche differences collectively stabilize the dynamics of experimental communities of serpentine annual plants. We used field-parameterized population models to develop a null expectation for community dynamics without the stabilizing effects of niche differences. The population growth rates predicted by this null model varied by several orders of magnitude between species, which is sufficient for rapid competitive exclusion. Moreover, after two generations of community change in the field, Shannon diversity was over 50 per cent greater in communities stabilized by niche differences relative to those exhibiting dynamics predicted by the null model. Finally, in an experiment manipulating species' relative abundances, population growth rates increased when species became rare—the demographic signature of niche differences. Our work thus provides strong evidence that species differences have a critical role in stabilizing species diversity.

For over a century, ecologists have explored the wide diversity of niche differences thought to stabilize coexistence⁸, exemplified by species' differences in rooting depth¹², the resources most limiting growth⁷ and interactions with specialist consumers^{13,14}. What unifies these differences is that they all cause species to limit themselves more than they limit their competitors¹⁵ (Fig. 1). Niche differences thus stabilize competitor dynamics by giving species higher per capita population growth rates when rare than when common (Fig. 1), and coexistence occurs when these stabilizing effects of niche differences overcome species differences in overall competitive ability. Although numerous studies have examined morphological, physiological and demographic differences between co-occurring species^{6–8,12}, the collective importance of those differences for the diversity observed in ecological communities is poorly understood¹¹. Ecologists have yet to determine whether species diversity is maintained by strong niche differences stabilizing the interactions of highly unequal competitors or, as suggested by the neutral theory⁹, whether niche differences are largely unimportant, only stabilizing the interactions of nearly equivalent competitors. More formally, these alternatives bracket a continuum of hypotheses concerning the importance of niches for diversity maintenance^{10,11}, one of the longest-standing problems in

ecology². Locating communities along this continuum is critical for understanding the fundamental stability of the diversity we observe in natural systems.

We evaluated the collective importance of niche differences by quantifying how rapidly species diversity decreases when the stabilizing effects of niche differences (advantages when rare and disadvantages when common) are eliminated from communities¹¹. The more important niche differences are for coexistence, the more rapidly inferior competitors are excluded when these differences are eliminated. Specifically, we used field-parameterized population models to predict the dynamics of an experimental community of annual plants under the condition that species lack niche differences^{15,16}. We

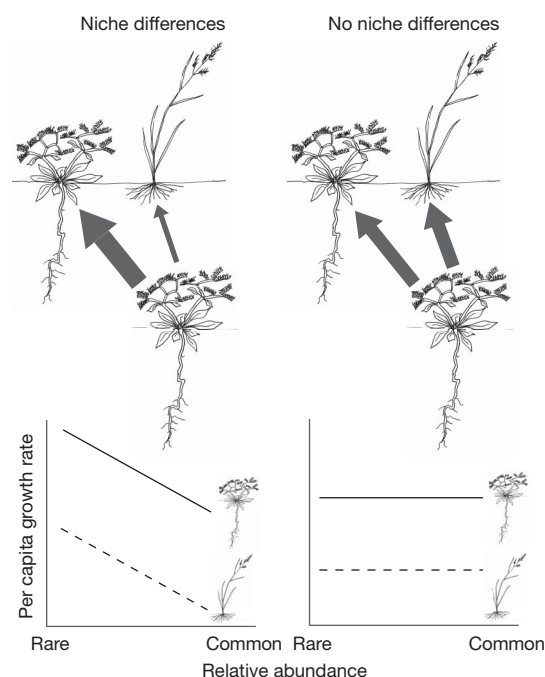


Figure 1 | How niche differences maintain diversity. Niche differences, including variation in rooting depth, cause species to limit individuals of their own species more than they limit competitors. This gives species greater per capita growth rates when they are rare and their competitors are common than when they are common and their competitors are rare. Such relationships stabilize coexistence by hindering competitors that reach high density and threaten other species with exclusion. With no niche differences, species limit themselves and their competitors equally, per capita growth rates do not change with species' relative abundances and variation between species reflects differences in fitness or competitive ability¹⁵. Arrow width represents the degree to which individuals limit one another.

¹Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, California 93106, USA. ²Department of Biology, University of Washington, Seattle, Washington 98195, USA.

then compared this null expectation to observed community dynamics in the field to quantify the impact of niche differences on coexistence. Finally, we tested for the demographic signature of these differences, namely greater per capita population growth rates when species are rare than when they are common (Fig. 1).

Our approach focused on experimentally assembled communities of annual plants on serpentine soils in California, USA. In the Mediterranean climate of our field site, annuals germinate in late autumn or early winter, and set seed in spring and summer. The system is well suited to our research aims because individuals are small, average 2,500 plants per square metre and reach high richness at small spatial scales¹⁷. The frequent co-occurrence of ten or more plant species per 0.0625 m² challenges niche-based theories of diversity maintenance. Most importantly, these annuals have relatively short and simple life cycles that can be reasonably described using the population models that form the basis of our approach.

We exploit the fact that niche differences influence coexistence by causing species to limit themselves more than they limit competitors (Fig. 1). We therefore predicted community dynamics without the stabilizing effects of niche differences as follows. We sowed ten replicate communities in the field, each with equal abundances of ten focal species that co-occur widely^{17,18} (Supplementary Table 1). We then parameterized commonly used annual-plant population models^{19–21} (Methods) with demographic rates measured in each community. Finally, we solved for each focal species' growth rate under the condition that communities are saturated with individuals and that species limit themselves and their competitors equally, as occurs without niche differences. Species differences in these growth rates reflect average competitive ability or fitness differences^{11,15} (Fig. 1).

Our theoretical approach predicts that without niche differences, species differ by several orders of magnitude in their per capita growth rates (Fig. 2a), which is sufficient for rapid competitive exclusion (Fig. 2b). For example, with 2007 demographic rates, the population size of *Navarretia atractyloides* was predicted to more than double per year, whereas that of the most inferior competitor species, *Micropus californicus*, was projected to decrease by 98% (Fig. 2a). We found similarly large variation among competitors with 2008 demographic rates, although in this wetter year the highest performing species was *Chorizanthe palmeri* (Fig. 2a). When these growth rates were averaged across years, *Salvia columbariae* had the highest predicted growth rate, 100 times greater than that of the most inferior species (Fig. 2a). Our theoretical approach is validated by our finding that after two generations of interaction in experimental communities, a species' relative abundance was correlated with its average growth rate predicted by the model (Spearman's rank correlation coefficient, 0.71; $P = 0.03$). The model can also approximate the rate of competitive exclusion without niche differences: communities would become 99.9% *Salvia* in less than 20 yr (Fig. 2b). This prediction emerged from simulations beginning with an equal abundance of all competitors. Each year, we randomly assigned 2007 or 2008 demographic rates, calculated the population growth rates (using equation (2) in Methods), and then updated species' relative abundances.

To quantify the influence of niche differences on coexistence in the field, we compared the dynamics of experimental communities stabilized by niche differences with that of communities experiencing the unstabilized population growth rates predicted by our null model. We established 20 replicate communities initially sown with an equal fraction of the ten competitors (by seed mass). Half of these communities were assigned the 'niche-removal treatment'. In each of these ten communities, we quantified model parameters and predicted population growth rates without niche differences (Methods). We then multiplied species' predicted growth rates by their seed numbers at the beginning of the growing season to determine the following year's seed composition (Supplementary Fig. 1). This process was repeated, each year incorporating year-specific demographic rates. By imposing population growth rates that were independent of species' commonness and rarity, this manipulation

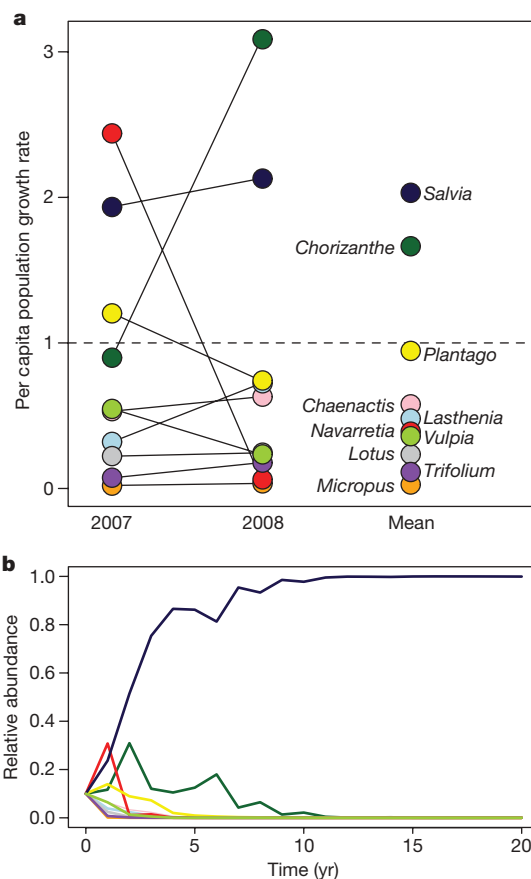


Figure 2 | Lack of competitive equivalence. **a**, Ten species' population growth rates (the number of individuals produced per individual, with species indicated by their genus) without the demographic influence of niche differences, for 2007 and 2008; the two-year geometric mean is also shown ($n = 10$). **b**, Theoretical projection of community dynamics without niche differences (the mean and median times to 99.9% dominance by *Salvia columbariae* are respectively 15.7 and 12 yr, based on 10,000 simulations). Colours correspond to species as in **a**.

removed the stabilizing effects of niche differences but retained species' differences in average competitive ability¹¹. We compared the dynamics with those in the remaining ten communities, used as controls, in which we replicated the seed-handling artefacts of the niche-removal treatment but retained the influence of niche differences (advantages when rare and disadvantages when common). In these communities, each year's seed composition was determined by species' measured seed production and the estimated seed bank carry-over (Supplementary Fig. 1).

After two generations of community change, Shannon diversity was 50% greater in communities stabilized by niche differences than in systems from which niche differences had been removed (treatment: $F_{1,36} = 51.2$, $P < 0.001$; year: $F_{1,36} = 48.6$, $P < 0.001$; treatment \times year: $F_{1,36} = 16.5$, $P < 0.001$ (from analysis of variance); Fig. 3). In both treatments, species composition shifted from an even abundance of all ten species to communities in which *Salvia columbariae* and *Plantago erecta* were more common. However, in the absence of niche differences the most common species, *Salvia columbariae*, became considerably more common, constituting almost 60% of 2008 community seed mass. Conversely, the seven rarest species constituted 35% of the community in the presence of niche differences, but only 8% in their absence. Given that niche differences influence coexistence by favouring species when they drop to low relative abundance^{11,15} (Fig. 1), our results qualitatively match the predictions of ecological theory: in the absence of niche differences, the common species become more common and the rare species

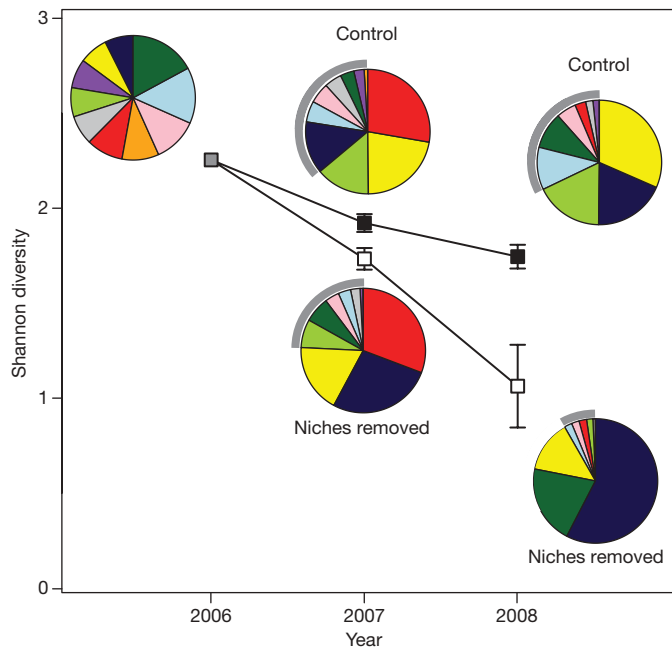


Figure 3 | Niche differences stabilize community dynamics. Two generations (2006–2007, 2007–2008) of change in the diversity and composition of communities stabilized by niche differences, versus those in which the demographic influence of niche differences was removed ($n = 10$). Pie charts show the average proportion of total community seed mass constituted by each focal species in each treatment and year. The grey arcs show the collective abundances of the seven rarest species. Species' relative abundances are not perfectly equal in the initial communities (2006) owing to differences in seed viability. Colours correspond to genus as in Fig. 2a and points show mean \pm s.e.

more rare (Fig. 3). Moreover, the observed changes in diversity in each treatment are too large to be explained by demographic stochasticity alone (Supplementary Fig. 2), as proposed by neutral theory.

Finally, we tested for the demographic signature of niche differences, namely species per capita population growth rates that increase as species become more rare¹¹ (Fig. 1). To accomplish this, we experimentally assembled serpentine annual communities and varied the relative abundance of each focal species from low to high. We then calculated a per capita population growth rate for each species by summing the number of seeds produced at the end of the growing season and the number of those carrying over in the seed bank.

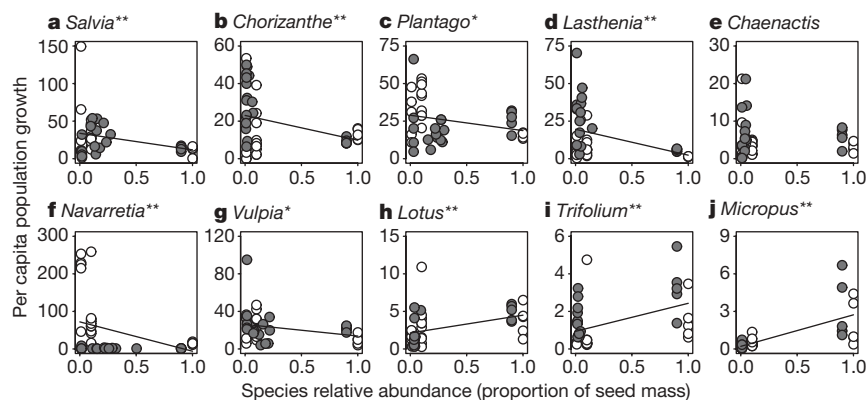


Figure 4 | Demographic effects of niche differences. The influence of a species' relative abundance in a community (commonness and rarity) on its population growth rate (the number of individuals produced per individual) in 2007 (open symbols) and 2008 (filled symbols). Species are ordered (a–j; referred to by their genus) by decreasing relative abundance in

Communities in which the influence of niche differences on dynamics has been removed (Fig. 3, 2008 pie chart for niche-removal treatment). The vertical-axis scale differs between plots. * $P < 0.10$, ** $P < 0.05$, from linear regression ($n = 40$).

Consistent with the expected influence of niche differences, the per capita population growth rates of the seven most abundant species decreased as each became increasingly common (although not significantly so for *Chaenactis glabriuscula*, whereas *Plantago erecta* and *Vulpia microstachys* had respective P values of 0.09 and 0.06) (Fig. 4a–g). By contrast, the three rarest species, presumably on their way to exclusion, showed positive relationships (Fig. 4h–j). These probably reflect intraspecific facilitative interactions (*Lotus wrangelianus* and *Trifolium willdenovii* are legumes) or the advantages these species experience when common and surrounded by other conspecific individuals of low competitive ability. More important than the number of species showing greater per capita population growth rates when rare than when common is the identity of those that did. *Salvia columbariae*, which dominated the communities from which niche differences had been removed, had a per capita growth rate that decreased by two-thirds as its relative abundance increased. The growth rate of *Chorizanthe palmeri*, the second most abundant in these communities, declined by one-half. Although the specific niche mechanisms responsible are unknown, *Salvia* can access a deeper resource base than all its competitors (Supplementary Table 1) and *Chorizanthe* grows several months later in the season than all but one of its competitors (Supplementary Table 1). These differences potentially stabilize their dynamics with the remainder of the community and contribute to patterns of relative abundance.

Our results support the hypothesis that niche differences strongly stabilize coexistence. However, our experiments probably miss niche mechanisms operating over larger spatial and longer temporal scales. For example, serpentine annual plants specialize on soil variation that occurs over tens of metres, which is not captured in our square-metre plots¹⁷. Similarly, species performing poorly in our experiment may germinate best under climatic conditions not experienced during the study. Given the spatial and temporal scale of our experiments, the importance of niche differences for coexistence proves unexpectedly strong.

Ecologists studying the maintenance of species diversity have traditionally examined individual coexistence mechanisms, such as resource partitioning^{3,7}, frequency-dependent enemy attack^{13,14} or the storage effect¹⁶. Our approach, by contrast, evaluates the collective importance of multiple niche mechanisms for coexistence. This is a critical distinction, because evidence for the latter uniquely justifies further study of individual niche mechanisms and bears on where natural communities fall along the continuum between classic niche theory and the neutral theory^{10,11}. Most importantly, our findings provide strong empirical support for the critical role niche differences have in stabilizing species diversity, one of the longest unresolved problems in ecology.

METHODS SUMMARY

Field work was conducted in a 500-m² area of serpentine habitat 30 km inland from Santa Barbara, USA. We parameterized the growth rates projected without niche differences (see equation (2) in Methods) in square-metre experimental communities. In autumn 2006, ten plots were sown with 15 g of seed per square metre, evenly divided between species. After recording germination, we thinned the plots to contain ≤ 10 individuals per species, from which we determined the seed production per germinant in the absence of competition. We measured seed bank survival by estimating seed viability using tetrazolium staining before and after a year of burial in nylon mesh bags. The year-specific growth rates for each species, calculated as described in full Methods, were averaged across plots to produce Fig. 2a.

We used the same plots to project communities forwards in the absence of niche differences (Supplementary Fig. 1). For each replicate, we multiplied each species' seed number at the beginning of the growing season by its theoretically projected growth rate over that season (calculated at season's end for each plot using plot-specific demographic rates). This product determined the seed mass added at the end of the growing season to a new plot adjacent to the previous year's community. Ten control communities experiencing the stabilizing effects of niche differences were of the same size, initial composition and total seed mass as the ten communities receiving treatment. Their reseeded amounts, however, were determined by species' actual seed production and seed bank carry-over. Shannon diversity ($-\sum_i p_i \ln p_i$) was calculated from each species' proportion, p_i , of total seed mass.

We quantified the relationship between species' per capita growth rate and their rarity and commonness in communities with 0.25-m² plots sown with 15 g of seed per square metre. Focal species frequency ranged from 1 to 100% of total seed mass, with replication concentrated at the extremes. The other nine competitors constituted the remaining seed mass in the communities.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 May; accepted 29 June 2009.

Published online 12 August 2009.

- Gause, G. F. *The Struggle for Existence* (Hafner, 1934).
- Hutchinson, G. E. The paradox of the plankton. *Am. Nat.* **95**, 137–145 (1961).
- Tilman, D. *Resource Competition and Community Structure* (Princeton Univ. Press, 1982).
- Volterra, V. Fluctuations in the abundance of a species considered mathematically. *Nature* **118**, 558–560 (1926).
- Lotka, A. J. The growth of mixed populations: two species competing for a common food supply. *J. Wash. Acad. Sci.* **22**, 461–469 (1932).
- MacArthur, R. H. *Geographical Ecology: Patterns in the Distribution of Species* (Harper & Row, 1972).
- McKane, R. B. et al. Resource-based niches provide a basis for plant species diversity and dominance in arctic tundra. *Nature* **415**, 68–71 (2002).
- Chase, J. M. & Leibold, M. A. *Ecological Niches: Linking Classical and Contemporary Approaches* (Univ. Chicago Press, 2003).
- Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, 2001).
- Leibold, M. A. & McPeck, M. A. Coexistence of the niche and neutral perspectives in community ecology. *Ecology* **87**, 1399–1410 (2006).
- Adler, P. B., HilleRisLambers, J. & Levine, J. M. A niche for neutrality. *Ecol. Lett.* **10**, 95–104 (2007).
- Cody, M. L. Niche theory and plant-growth form. *Vegetatio* **97**, 39–55 (1991).
- Chesson, P. & Kuang, J. J. The interaction between predation and competition. *Nature* **456**, 235–238 (2008).
- Hille Ris Lambers, J., Clark, J. S. & Beckage, B. Density-dependent mortality and the latitudinal gradient in species diversity. *Nature* **417**, 732–735 (2002).
- Chesson, P. Mechanisms of maintenance of species diversity. *Annu. Rev. Ecol. Syst.* **31**, 343–366 (2000).
- Adler, P. B., HilleRisLambers, J., Kyriakidis, P. C., Guan, Q. & Levine, J. M. Climate variability has a stabilizing effect on coexistence of prairie grasses. *Proc. Natl Acad. Sci. USA* **103**, 12793–12798 (2006).
- Harrison, S. Local and regional diversity in a patchy landscape: native, alien, and endemic herbs on serpentine. *Ecology* **80**, 70–80 (1999).
- Hobbs, R. J., Yates, S. & Mooney, H. A. Long-term data reveal complex dynamics in grassland in relation to climate and disturbance. *Ecol. Monogr.* **77**, 545–568 (2007).
- Watkinson, A. R. Density-dependence in single-species populations of plants. *J. Theor. Biol.* **83**, 345–357 (1980).
- Pacala, S. W. Neighborhood models of plant population dynamics. 4. Single-species and multispecies models of annuals with dormant seeds. *Am. Nat.* **128**, 859–878 (1986).
- Levine, J. M. & Rees, M. Effects of temporal variability on rare plant persistence in annual systems. *Am. Nat.* **164**, 350–363 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by US National Science Foundation grants 0743365 and 0743183, and a fellowship from the David and Lucile Packard Foundation. P. Adler, C. Briggs, B. Cardinale, P. Chesson, M. Levine, D. Murrell and the Levine and HilleRisLambers laboratories provided comments on the manuscript. C. Cowan, R. Harris and C. Peters conducted the field work.

Author Contributions J.M.L. and J.H. jointly conducted the project, analysed the data and prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.M.L. (levine@lifesci.ucsb.edu) or J.H. (jhrl@u.washington.edu).

METHODS

Study system. We conducted our study in serpentine habitat at the University of California Sedgwick Reserve in Santa Barbara County, USA. The climate is Mediterranean with cool, wet winters and hot, dry summers. Annual precipitation at the reserve was 19.8 cm in 2006–2007, and 40.1 cm in 2007–2008 (38 cm is the 50-yr average). The site is dominated by annual plants, which germinate in late autumn or early winter and set seed in spring and summer. Our experimental communities were assembled in areas cleared of all vegetation (mostly exotic annual grasses) and subsequently weeded to ensure our direct control over community composition. Weed matting lay between the experimental communities.

All seed for experimental communities was locally collected, primarily from the rockier portions of the habitat where our focal species still dominate. Locating experimental communities in these rockier habitats was not feasible, owing to their limited extent and the pre-existing seed bank of the focal species. Our experiment focused on ten native annual plants (Supplementary Table 1) covering a range of natural abundances.

Theoretical approach. To project species' population growth rates without the demographic influence of niche differences, we first defined a model that could reasonably describe competitor dynamics in our annual communities. We then empirically obtained the demographic rates necessary for calculating population growth rates in the hypothetical case in which species limit themselves and their competitors equally. We began with the following well-studied two-species annual-plant model^{19–21}. Maximum-likelihood analyses showed that, relative to seven other candidate models, this model best described how seed production changed with density in the experimental communities (Supplementary Table 2). The population growth rate for species i competing with species j is modelled as follows:

$$\frac{N_{i,t+1}}{N_{i,t}} = s_i(1 - g_i) + \frac{\lambda_i g_i}{1 + \alpha_{ii} g_i N_{i,t} + \alpha_{ij} g_j N_{j,t}} \quad (1)$$

Here $N_{i,t}$ is the number of seeds of species i at the beginning of the growing season of year t before germination. The first term of the sum describes the carry-over of seeds in the seed bank, a function of g_i , the fraction of germinating seeds, and s_i , the annual survival of ungerminated seed in the soil. The second term describes population growth due to germination and eventual seed production: λ_i is the number of viable seeds produced per germinated individual in the absence of competition, and α_{ij} is a competition coefficient describing the effect of a germinated individual of species j on the seed produced per germinant of species i (these differ from the relative α coefficients of the Lotka–Volterra equations¹⁵). Importantly, the terms involving the competition coefficients are phenomenological and represent all processes by which individuals limit one another, including resource competition and interactions with shared consumers and pathogens¹³. Interchanging all i and j subscripts gives the model for species j .

To approximate the growth rate of species i without the demographic influence of niche differences, we imposed two conditions. First, we forced species to limit themselves and their competitors equally by setting the per capita effects of each species on their own growth to equal their effects on competitors ($\alpha_{ij} = \alpha_{ji}$ and $\alpha_{ii} = \alpha_{jj}$). Second, we assumed that for any density of species i , the abundance of species j is equilibrated²², which in effect fills the community with individuals. Under these two conditions, we obtained the following growth rate (see Supplementary Methods for details and an alternative approach):

$$\frac{N_{i,t+1}}{N_{i,t}} = s_i(1 - g_i) + \lambda_i g_i \left[\frac{1 - s_j(1 - g_j)}{\lambda_j g_j} \right] \quad (2)$$

This per capita growth rate is independent of species' relative abundances, as expected in a fully saturated community (the second condition) without niche differences (Fig. 1). Moreover, in these high-density competitive systems, species' germination, survival and low-density fecundity, all of which we measure in our experimental communities (Methods Summary), determine dominance.

Equation (2) separates the demographic rates for species i from those of its competitor, species j , which are in square brackets (see ref. 23 for interpretation of this term). Because our experimental communities are composed of ten rather than two competitors, we averaged the bracketed term for each of the nine

competitors faced by species i and weighted this average by the competitors' relative abundances (relative abundances were initially equal). For our ten-species community, using equation (2) to project growth rates without niche differences meant that we forced species i to equally limit itself and the nine competitors it faces, and these competitors collectively to limit themselves and species i to the same extent. All predicted growth rates were scaled such that total seed mass in a community did not change between years.

Obtaining parameters for species per capita growth rates without niche differences. We measured the demographic parameters in equation (2) for each of the ten competitors in each growing season (2006–2007, 2007–2008). We measured the germination rate in ten circular plots sown with a mixture of the ten focal species (15 g of seed per square metre). Plots were 0.5 m² in 2006–2007 and were enlarged to 1 m² in 2007–2008 owing to greater seed availability. Germination was recorded by placing coloured plastic toothpicks adjacent to each germinant in multiple visits to each plot over the winter. In 2007–2008, we measured the number of seeds produced per germinant in the absence of competition (λ_i) by thinning the ten plots (after germination) down to no more than ten individuals per species. We harvested all seed from those plants to determine the seed production per germinant, and corrected that number for seed viability. In 2006–2007, we measured λ_i by thinning down five 0.0625-m² plots per species and using the same methods as described for 2007–2008.

We measured seed bank survival by estimating seed viability before and after a year of burial in ten nylon mesh bags per species. We measured seed viability by placing seeds on wetted germination paper in a cold room (15 °C) for five days, and then stored them at room temperature (22 °C) until germination ceased. We determined the viability of ungerminated seeds by immersing them in gibberellic acid and, 24 h later, cutting and staining the seeds with tetrazolium²⁴. Those that stained viable were added to the number of germinants to yield total viability.

Measuring the relationship between population growth and species commonness and rarity. In autumn 2006 and autumn 2007, we established 110 circular plots, each 0.25 m² in area. All plots were sown at a density of 15 g of seed per square metre, and were relocated each year to prevent uncontrolled seed bank carry-over. Ten of the plots were 'natural dynamics plots' sown with an equal proportion of the ten competitors in 2006. In autumn 2007, they were sown at a relative abundance matching that found at the end of the 2006–2007 growing season. The remaining 100 plots were equally divided between low-frequency and high-frequency plots for each species. Specifically, we sowed five low-frequency plots per species in which 1% of the total seed mass belonged to the focal species; the remaining 99% of the seed mass consisted of the nine other competitors, with their relative abundances matching those in the natural dynamics plots. Each focal species was also assigned to five high-frequency plots, in which it was sown at 100% of total seed mass in 2006 and 90% of total seed mass in 2007. Owing to limited seed in the first year of the project (2006), the 1% plots and some of the high-frequency plots were 0.0625 m² in size that year.

We estimated species per capita population growth rate, $N_{i,t+1}/N_{i,t}$, in each community using the following equation:

$$\frac{N_{i,t+1}}{N_{i,t}} = s_i(1 - g_i) + F_i g_i$$

Here s_i and g_i are seed survival and germination, measured as described in the previous section, and F_i is the number of viable seeds produced per germinant, implicitly incorporating all intra- and interspecific interactions that occur over the growing season. We measured F_i for each focal species by harvesting all of its seeds as they ripened in a plot and then dividing the total seed number by the number of germinants. Finally, we corrected these values for seed viability, measured as described in the previous section.

22. Levine, J. M., Adler, P. B., & HilleRisLambers, J. On testing the role of niche differences in stabilizing coexistence. *Funct. Ecol.* **22**, 934–936 (2008).
23. Kuang, J. J. & Chesson, P. Coexistence of annual plants: generalist seed predation weakens the storage effect. *Ecology* **90**, 170–182 (2009).
24. Baskin, C. C. & Baskin, J. M. *Seeds: Ecology, Biogeography and Evolution of Dormancy and Germination* 19 (Academic, 2001).

LETTERS

Photosystem I gene cassettes are present in marine virus genomes

Itai Sharon^{1,2*}, Ariella Alperovitch^{1*}, Forest Rohwer^{4,5}, Matthew Haynes⁴, Fabian Glaser³, Nof Atamna-Ismaeel¹, Ron Y. Pinter², Frédéric Partensky⁶, Eugene V. Koonin⁷, Yuri I. Wolf⁷, Nathan Nelson⁸ & Oded Béjà¹

Cyanobacteria of the *Synechococcus* and *Prochlorococcus* genera are important contributors to photosynthetic productivity in the open oceans^{1–3}. Recently, core photosystem II (PSII) genes were identified in cyanophages and proposed to function in photosynthesis and in increasing viral fitness by supplementing the host production of these proteins^{4–7}. Here we show evidence for the presence of photosystem I (PSI) genes in the genomes of viruses that infect these marine cyanobacteria, using pre-existing metagenomic data from the global ocean sampling expedition⁸ as well as from viral biomes⁹. The seven cyanobacterial core PSI genes identified in this study, *psaA*, *B*, *C*, *D*, *E*, *K* and a unique *J* and *F* fusion, form a cluster in cyanophage genomes, suggestive of selection for a distinct function in the virus life cycle. The existence of this PSI cluster was confirmed with overlapping and long polymerase chain reaction on environmental DNA from the Northern Line Islands. Potentially, the seven proteins encoded by the viral genes are sufficient to form an intact monomeric PSI complex. Projection of viral predicted peptides on the cyanobacterial PSI crystal structure¹⁰ suggested that the viral–PSI components might provide a unique way of funnelling reducing power from respiratory and other electron transfer chains to the PSI.

Bacteriophages have the ability to manipulate the life histories and evolution of their hosts¹¹ and evolved many adaptation and defence mechanisms for efficient survival and multiplication. Most of these involve manipulation of the host DNA, as well as the incorporation, into the phage genomes, of bacterial genes that encode proteins with a potential to facilitate bacteriophage reproduction¹². Recently, it was discovered that marine cyanophages (bacteriophages that infect cyanobacteria) carry photosynthetic genes, and it was suggested that these genes increase phage fitness^{4–7}. Cyanobacterial photosynthetic membranes contain two photosystems, of which PSII mediates the transfer of electrons from water, the initial electron donor, to the plastoquinone pool, whereas PSI mediates electron transfer from plastocyanin to ferredoxin, thereby generating reducing power needed for CO₂ fixation in the form of NADPH. Although PSII is known to be sensitive to photodamage, PSI is considered to be a more stable complex. The PSII gene *psbA* coding for the labile D1 protein is readily detected in various cultured and environmental cyanophages infecting *Prochlorococcus* and *Synechococcus*^{4,6,13,14}. Furthermore, other photosynthesis genes encoding the PSII D2 protein^{4,6}, high-light inducible proteins, pigment biosynthesis proteins (Ho1, PebA and PcyA), or the photosynthetic electron transport proteins plastocyanin (PetE) and ferredoxin (PetF) were also identified in several cyanophage genomes^{6,15}.

To assess the possible presence of other photosynthesis-related genes in viruses, we set up a designated search scheme for publicly

available metagenomic data. Initially we searched for the cyanobacterial PSI gene *psaA*. Together with PsaB, the PsaA protein forms the heterodimeric core of PSI that binds the primary electron donor P700, formed by a special chlorophyll pair¹⁰. Using tBLASTx, different *Synechococcus* and *Prochlorococcus* *psaA* gene sequences were used as queries against the global ocean sampling (GOS) expedition⁸ data set.

We detected 574 *psaA*-containing GOS scaffolds. These were further screened to identify those that were likely to originate from viruses using tBLASTx against refseq_viral, a database that contains all known viral genomes. This procedure reduced the number of suspected scaffolds to five. The PsaA homologues encoded by these sequences showed only 65–75% identity to *Prochlorococcus* or marine *Synechococcus* PsaA proteins. On a maximum-likelihood tree, four of these proteins clustered together on a well-supported branch related to *Prochlorococcus* PsaA, whereas the fifth sequence (JCVI_SCAF_1096628008692) was retrieved near the base of the *Synechococcus* branch (Fig. 1). Because the GOS general scaffold assembly represents reads that come from different GOS sample sites or from different clones and hence are chimaerical by definition, we restricted all further analysis to sequences assembled from single clone reads only. Analysis of the GOS clones containing the modified *psaA* genes confirmed their viral origin (probably cyanophages of the *Myoviridae* family), as indicated by the presence, in the vicinity of *psaA*, of typical viral genes, such as *nrdA* and *B* (that encode the α 2 and β 2 subunits of viral ribonucleoside diphosphate reductase, respectively) or the T4-like neck gp13 protein gene (Fig. 2). In addition to *psaA*, these clones contained clusters of PSI genes, including *psaB*, *psaC*, and a unique fused version of the *psaF* and *psaJ* genes (*psaJF*). An analysis of the GOS data sets with other PSI peptides as baits showed the presence of several other PSI clusters also containing *psaE*, *psaK* and *psaD* genes (see distribution in the different GOS sites in Supplementary Table 1). Like with PsaA, phylogenies made with these extra PSI protein sequences showed that they were all clustered at a distance from the homologous proteins of *Prochlorococcus* and *Synechococcus*, with the exception of PsaC and PsaD from GOS clone 1061008099984 (hereafter described as clone 9984; a clone used to build the previously mentioned scaffold JCVI_SCAF_1096628008692), which were retrieved closer to corresponding cyanobacterial sequences than to other viral sequences (Supplementary Fig. 1). Examining the *Prochlorococcus* and *Synechococcus* genome arrangements (Fig. 2, middle panel) or gene-pairs frequency modelling showed that the organization observed in most viral clones, *psaJF*-*C*-*A*-*B*-*K*-*E*-*D*, differs from that observed in these cultured cyanobacterial genomes and in most other (probably cyanobacteria-derived) GOS sequences (Fig. 3).

¹Faculty of Biology, ²Faculty of Computer Science, ³Bioinformatics Knowledge Unit, Lorry I. Lokey Interdisciplinary Center for Life Sciences and Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel. ⁴Department of Biology, ⁵Center for Microbial Sciences, San Diego State University, San Diego, California 92182, USA. ⁶CNRS and UPMC-Université Paris 6 (UMR 7144), Station Biologique, 29682 Roscoff, France. ⁷National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA. ⁸Department of Biochemistry, George S. Wise Faculty of Life Sciences, Daniella Rich Institute for Structural Biology, Tel Aviv University, Tel Aviv 69978, Israel.

*These authors contributed equally to this work.

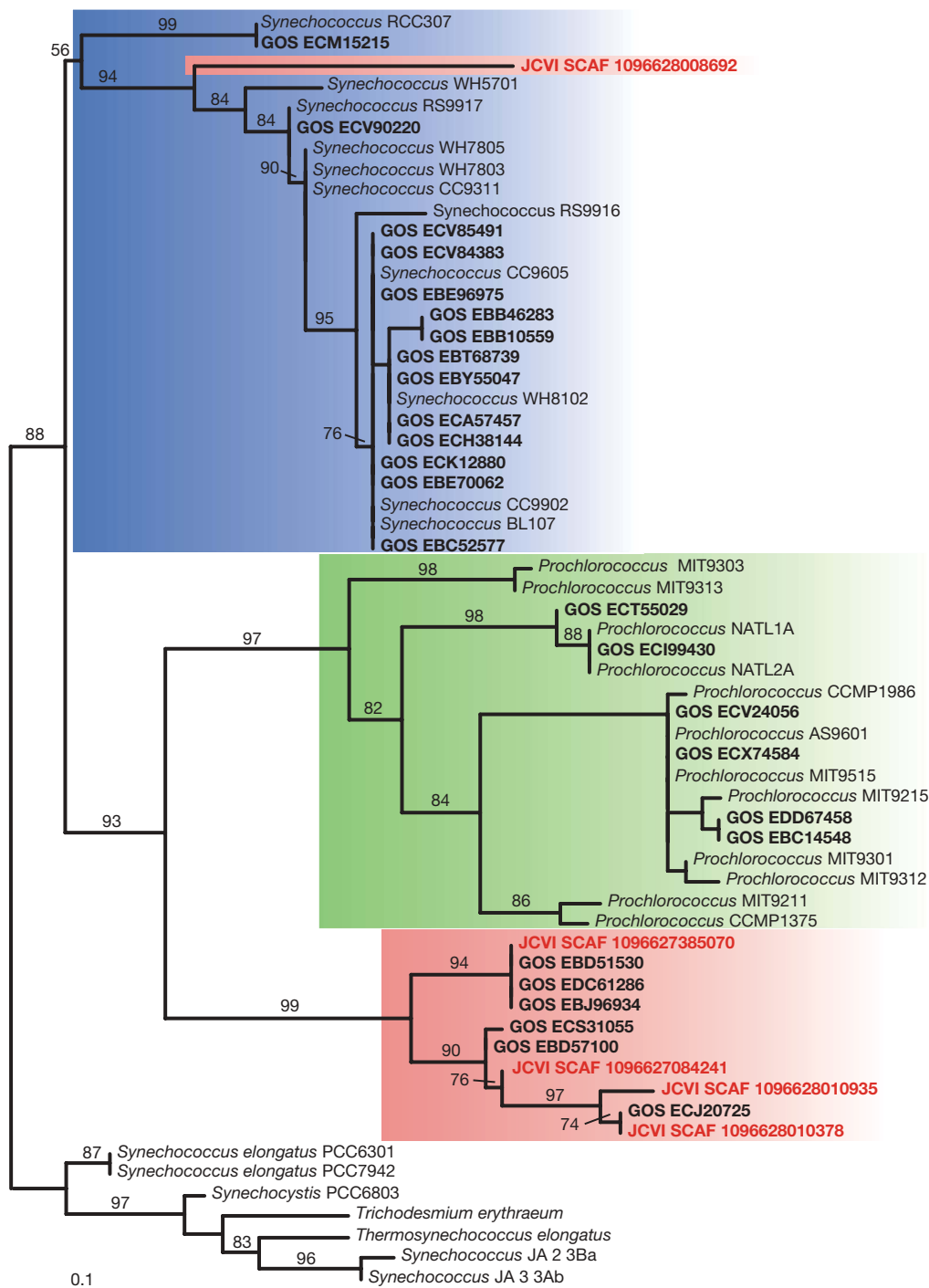


Figure 1 | A maximum-likelihood phylogenetic tree of *psaA*-deduced amino acid sequences obtained from the GOS expedition. *PsaA* sequences from the 27 fully sequenced and annotated *Synechococcus* (blue background) and *Prochlorococcus* (green background) genomes are shown. Sequences from the GOS expedition are shown in bold, and sequences from the original five scaffolds obtained in this study are indicated in red. For clarity, the tree shows only a subset of the 583 partial *PsaA* sequences found in the GOS data set. The tree is on the basis of an alignment of 94 shared amino acids. See Supplementary Methods for description of tree construction.

The PSI genes found on clone 9984 (represented by GOS reads 1095964115098 and 1095975140994 in Fig. 2) had a different order (*psaD-C-A*) than on the other clones, consistently with their distinct positions in phylogenetic trees (Supplementary Fig. 1).

To validate the viral origin of these genes and their unique cluster organization, data obtained from the GOS project were cross-referenced with recently released 454 pyrosequencing metagenomic sequences obtained from a variety of marine and non-marine viral and microbial biomes data sets⁹. This was a critical step in increasing the credibility of the results because the two approaches each introduce different biases¹⁶. The various viral-suspected PSI GOS clones identified were used to recruit reads from these different data sets. Marine virome fragments were readily recruited to all of the viral GOS clones regions, whereas virome or microbiome fragments coming from other environments were scarcely recruited (Table 1), with a much lower identity (Fig. 4a),

further supporting a marine viral origin for the PSI clones. The overall coverage measure of viromes and microbiomes to all different GOS clones containing PSI genes (Fig. 4b) clearly points to two distinguished populations, one from bacteria (cyanobacteria) and one from viruses (phages). Except for clone 9984, all our identified viral clones are falling in the viral population. Furthermore, marine virome fragments were also recruited to regions between the photosynthesis genes, linking neighbour genes in the observed viral cassette (Fig. 4a and Supplementary Table 2)—an observation that supports the gene cluster organization observed on the GOS clones.

To validate the juxtaposition of the genes in the identified viral-PSI gene clusters, DNA from the Northern Line Islands marine virome¹⁷ was used to perform 'continuous' overlapping and long PCR with primers assigned to the different genes (Supplementary Table 5). The results of the 'continuous' overlapping PCR (Fig. 2, bottom

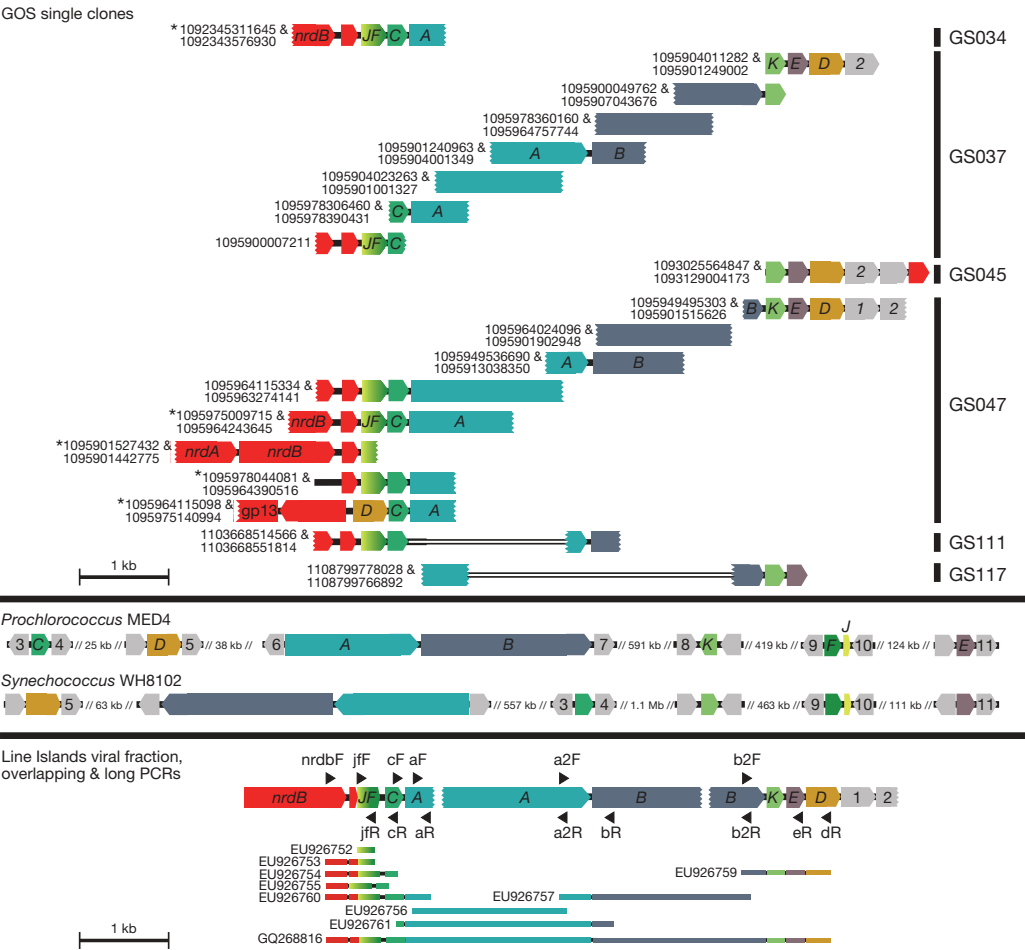


Figure 2 | Schematic physical maps of selected viral-suspected GOS clones (top), *Prochlorococcus* and *Synechococcus* genomes (middle) and environmental PCR products containing PSI genes (bottom). Red arrows represent ORFs with predicted viral origin, and grey arrows represent unknown ORFs. Capital letters represent the corresponding PSI core genes. Gaps shown in Indian Ocean GOS clones (stations GS111 and GS117) are the result of regions that were not covered by the end-reads owing to the size of these clones (5 kb). Primer positions on GOS clones are indicated by triangles, and thick coloured lines denote PCR products.

panel) and the amplification of a ~6.2-kilobase (kb) long PCR amplicon (Supplementary Fig. 2 and Fig. 2, bottom panel) spanning the entire PSI cassette and including the viral *nrdB* gene, show that the different genes in the cluster *nrdB-hyp-psaJF-C-A-B-K-E-D* (and also a new arrangement *nrdB-psaJF-C*; GenBank accession EU926755) are physically linked and exist as one photosynthetic cluster.

Although the data presented here are derived from environmental genomic data sets (non-continuous data), and therefore the lack of genes is not a proof of absence, it is notable that the PSI genes *psaI*,

psaL and *psaM* were not found in the viral *psa* gene cassettes. The *psaM* gene is naturally absent from plants¹⁸ and its inactivation in cyanobacteria shows that it is mainly required for the formation of stable PSI trimers¹⁹. Similarly, targeted inactivation of cyanobacterial *psaL* produces functional PSI complexes unable to form trimers, whereas *PsaI* is mostly required for stabilizing *PsaL*²⁰. Therefore, these three proteins are mainly involved in the trimer formation of cyanobacterial PSI, and their potential absence from the viral clone might indicate the formation of a monomeric PSI complex as in

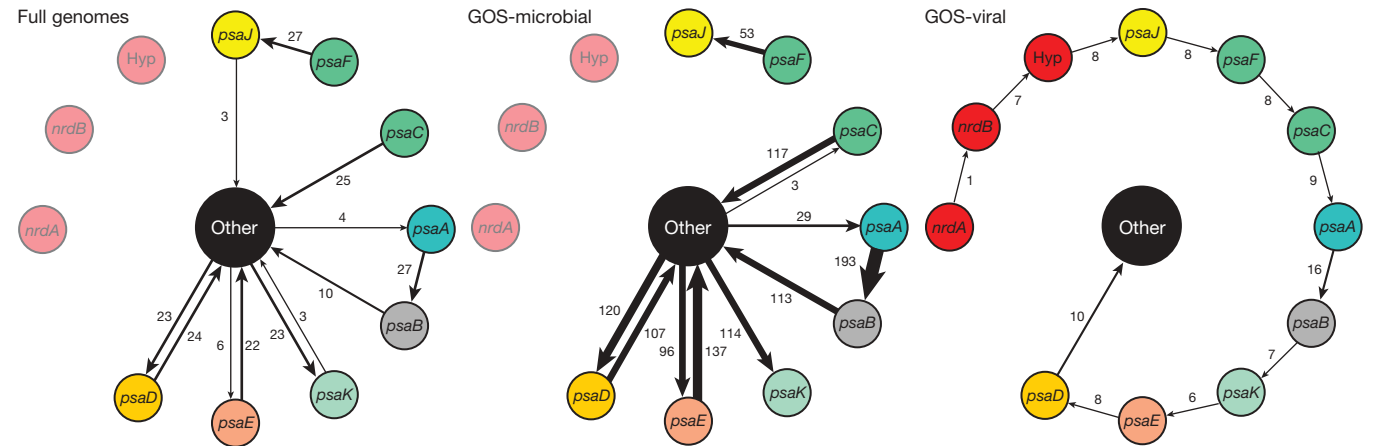


Figure 3 | Distribution of neighbouring genes involving at least one PSI gene. Each arrow connects neighbouring genes, and its thickness represents the number of pairs found in *Synechococcus* and *Prochlorococcus* genomes (left gene-circle), microbial sequences from the GOS metagenome (middle gene-circle), and viral sequences from the GOS metagenome (right

gene-circle). Note the uninterrupted clustering of PSI genes in phage genomes that contrasts the scattered arrangement of these genes in cyanobacterial genomes (in both cultures and GOS). Gene connections observed only once are not shown.

Table 1 | Number of different biome reads recruited to GOS-suspected viral-PSI clones

Type	Microbial metagenomes			Viral metagenomes		
	85%	90%	95%	85%	90%	95%
Coral	0	0	0	0	0	0
Fish	0	0	0	0	0	0
Freshwater	2	0	0	1	1	1
Hypersaline	1	0	0	0	0	0
Marine	1	0	0	207	144	91
Microbialites	0	0	0	0	0	0
Terrestrial	0	0	0	0	0	0

plants²¹ and not a trimeric complex as in cyanobacteria²². All genetic information required to form this putative minimal, monomeric PSI is clustered onto a very small cyanophage genome fragment (~5.9 kb). To our knowledge, gene clusters encoding all the components of a photosystem from an oxygenic phototroph have not been previously reported, and neither have there been reports on cyanobacterial PSI genes outside a cyanobacterial chromosome.

The potential structural consequences of assembling the phage proteins into the PSI complex were modelled in relation to the 2.5 Å structure of PSI from the cyanobacterium *Thermosynechococcus*

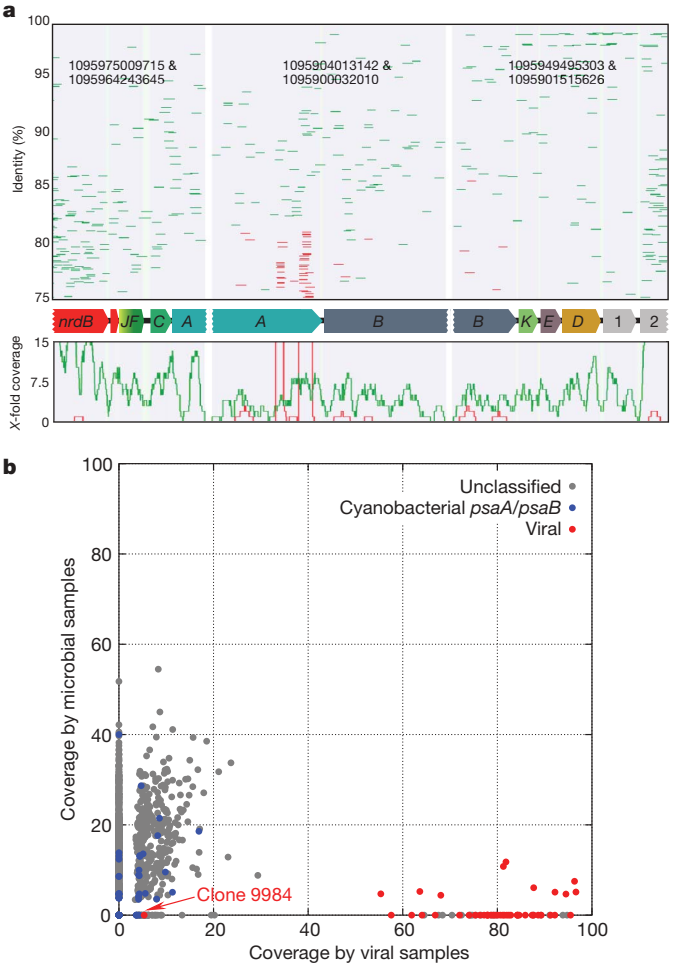


Figure 4 | Recruitment of GOS clones carrying PSI genes with Northern Line Islands biomes. **a**, Recruitment of three GOS physical clones carrying suspected viral-PSI genes by Northern Line Islands virome (green) and microbiome (red) reads. The top panel shows recruitment at 75–100% identity, and the bottom panel shows the fold-coverage by these reads. Accession numbers of the GOS reads used are presented above each clone (JCVC_READ_#). **b**, Recruitment coverage of GOS single clones carrying PSI genes with Northern Line Islands biomes (viromes (x-axis) and microbiomes (y-axis)). Coverage is defined as the percentage of GOS clone length covered by at least one recruited read.

*elongatus*¹⁰. We modelled the PsaJF fusion protein (in which the carboxy terminus of PsaJ is fused to the amino terminus of PsaF) at the position of subunits J and F of PSI. Figure 5 shows that the viral PsaJF fusion protein fits perfectly at the position of subunits J and F in the PSI structure. The only prominent change was the absence of the N terminus of subunit F, which is responsible for the specific binding of the natural electron donor (plastocyanin) of PSI²³. In chloroplasts of green algae and plants, this part of subunit F is elongated, resulting in higher affinity of plastocyanin to the chloroplast PSI^{18,23,24}. Although both plastocyanin and cytochrome *c*₆ are capable of donating electrons to PSI²⁴ in *Chlamydomonas reinhardtii*, this site in higher plants is specific for plastocyanin²⁵. However, the electron donation to PSI is not at all promiscuous, and several soluble cytochromes, including the respiratory cytochrome *c*, fail to donate electrons to PSI²⁶. We propose that the replacement of PsaJ and PsaF with the viral PsaJF fusion protein enables electron donation through extra electron carriers, including cytochromes that usually function as electron donors to cytochrome oxidase.

The mechanistic consequence of a less selective electron donation to PSI might be the possibility of sharing reducing power generated by the respiratory chain with the photosynthetic electron transport chain. A similar phenomenon, called chloro-respiration, detected in both cyanobacteria and chloroplasts, was attributed to the plastid terminal plastoquinone oxidase (PTOX)²⁷. The electron mediator in

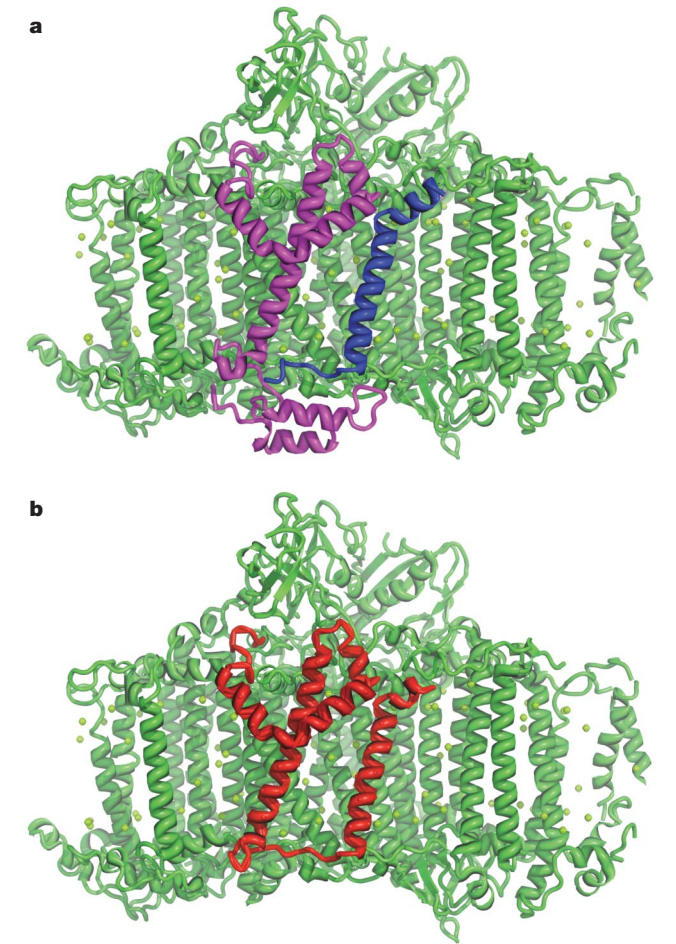


Figure 5 | Structural consequences of assembling the viral fusion protein PsaJF into PSI. **a**, The structure of *T. elongatus* PSI (subunits) was illustrated by PyMOL (<http://pymol.sourceforge.net/>) using a PSI monomer (adopted from Protein Data Bank (PDB) accession 1jb0). PsaF is in magenta, PsaJ is in blue, and all of the other subunits are in green. **b**, A model for the structure of the viral PsaJF fusion protein (red) substituting the original PsaF and PsaJ subunits.

this process is plastoquinone, which shuttles between the respiratory-like chain and the chloroplast *b₆f* complex²⁷. After phage infection and the incorporation of the phage gene products into PSI, the function of electron mediation could be carried out by a soluble cytochrome. Moreover, the phage might boost the amount of PSI to lead the infected cyanobacterial cells towards a cyclic photosynthesis for the generation of ATP in expense for the production of reducing power for CO₂ fixation. The PSI levels are notably low in both oceanic *Synechococcus*²⁸ and in *Prochlorococcus*²⁹, possibly as a result of adaptation to low iron levels, and it was recently proposed that a compensatory mechanism might exist, involving alternative electron flow to O₂ (ref. 28).

The phage PSI gene fusion *psaJF* described here is, to our knowledge, the first example of a phage gene innovation that involves structural membrane proteins. Modification towards a new function of existing cyanobacterial proteins by their phages was recently demonstrated for the divergent phage *PebA* homologue³⁰ (renamed *PebS* (phycocerythrobilin synthase)). The phage *PebS* single-handedly catalyses a reaction for which uninfected host cells require two consecutive enzymes, *PebA* and *PebB*. Considering these findings and our calculations that suggest a high likelihood of gene cluster formation in phage genomes (see Supplementary Information), the oceanic virome could be an almost unlimited source of naturally bioengineered gene cassettes.

METHODS SUMMARY

Collecting GOS-PSI clones. The following steps were taken to identify viral and non-viral PSI clones in GOS: (1) tBLASTx searches, with e-value threshold of 10⁻²⁰, of *psaA*, *psaB*, *psaC*, *psaD*, *psaE*, *psaF*, *psaJ* and *psaK* probes against the data set of GOS scaffolds. This step yielded 1,167 scaffolds. (2) Identify all reads composing the scaffolds found in the previous stage and their division into clones. Overall, 3,758 reads from 2,147 clones were found (536 single-read and 1,611 pair-end clones). (3) 'In-clone assembly', the reads of each pair-end clone were aligned (bl2seq) and assembled; 50 Ns were added between non-overlapping reads. (4) Annotation, an iterative procedure was used for gene discovery and annotation: at each iteration all clones were BLASTed against nr (e-value threshold = 10), first hit for each clone was saved and the clone's segment in the alignment was replaced with Ns. For each clone, the process halted when no new hits were found. (5) All clones with no PSI hit were removed. Overall we were left with 1,585 GOS clones carrying at least one PSI gene.

Collecting preliminary set of viral-PSI sequences in GOS. To find candidate viral-PSI sequences in GOS we have used the following two-step method: first, identify all GOS sequences (scaffolds or clones) containing *psaA* genes (see earlier); and second, identify viral genes on *psaA*-containing sequences. In the second step, all *psaA*-containing sequences were blasted (tBLASTx) against the refseq-viral database (again, with an e-value threshold of 10⁻²⁰). The initial scan revealed five scaffolds containing both *psaA* and viral genes. These scaffolds were annotated (BLASTx against the nr database) and found to contain both viral genes such as *nrdA* and *nrdB*, as well as PSI genes such as *psaA*, *psaC*, *psaD* and a fusion of *psaJ* and *psaF*.

For details on recruitments against 454 databases, gene organization analysis, abundance measures, estimation of the number of recombination events, and PCR conditions, see Supplementary Methods.

Received 14 June; accepted 9 July 2009.

Published online 26 August 2009.

- Li, W. K. W., Zohary, T., Yacobi, Y. Z. & Wood, A. M. Ultraphytoplankton in the eastern Mediterranean Sea: towards deriving phytoplankton biomass from flow cytometric measurements of abundance, fluorescence and light scatter. *Mar. Ecol. Prog. Ser.* **102**, 79–87 (1993).
- Liu, H., Nolla, H. A. & Campbell, L. *Prochlorococcus* growth rate and contribution to primary production in the equatorial and subtropical North Pacific Ocean. *Aquat. Microb. Ecol.* **12**, 39–47 (1997).
- Partensky, F., Hess, W. R. & Vaulot, D. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol. Mol. Biol. Rev.* **63**, 106–127 (1999).
- Mann, N. H., Cook, A., Millard, A., Bailey, S. & Clokie, M. Bacterial photosynthesis genes in a virus. *Nature* **424**, 741 (2003).
- Millard, A., Clokie, M. R. J., Shub, D. A. & Mann, N. H. Genetic organization of the *psbAD* region in phages infecting marine *Synechococcus* strains. *Proc. Natl Acad. Sci. USA* **101**, 11007–11012 (2004).
- Lindell, D. et al. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc. Natl Acad. Sci. USA* **101**, 11013–11018 (2004).

- Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M. & Chisholm, S. W. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**, 86–89 (2005).
- Rusch, D. B. et al. The *Sorcerer II* Global Ocean Sampling expedition: northwest Atlantic through the eastern tropical Pacific. *PLoS Biol.* **5**, e77 (2007).
- Dinsdale, E. A. et al. Functional metagenomic profiling of nine biomes. *Nature* **452**, 629–632 (2008).
- Jordan, P. et al. Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature* **411**, 909–917 (2001).
- Rohwer, F. & Thurber, R. V. Viruses manipulate the marine environment. *Nature* **459**, 207–212 (2009).
- Brown, N. F., Wickham, M. E., Coombes, B. K. & Finlay, B. B. Crossing the line: selection and evolution of virulence traits. *PLoS Pathog.* **2**, e42 (2006).
- Sullivan, M. B. et al. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol.* **4**, e234 (2006).
- Sharon, I. et al. Viral photosynthetic reaction center genes and transcripts in the marine environment. *ISME J.* **1**, 492–501 (2007).
- Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F. & Chisholm, S. W. Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. *PLoS Biol.* **3**, e144 (2005).
- Harismendy, O. et al. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol.* **10**, R32 (2009).
- Dinsdale, E. A. et al. Microbial ecology of four coral atolls in the northern line islands. *PLoS One* **3**, e1584 (2008).
- Ben-Shem, A., Frolov, F. & Nelson, N. The crystal structure of plant photosystem I. *Nature* **426**, 630–635 (2003).
- Naithani, S., Hou, J. M. & Chitnis, P. R. Targeted inactivation of the *psaK1*, *psaK2* and *psaM* genes encoding subunits of Photosystem I in the cyanobacterium *Synechocystis* sp. PCC 6803. *Photosynth. Res.* **63**, 225–236 (2000).
- Xu, Q. et al. Mutational analysis of photosystem I polypeptides in the cyanobacterium *Synechocystis* sp. PCC 6803. Targeted inactivation of *psaL* reveals the function of *psaL* in the structural organization of *psaL*. *J. Biol. Chem.* **270**, 16243–16250 (1995).
- Kouril, R., van Oosterwijk, N., Yakushevskaya, A. E. & Boekema, E. J. Photosystem I: a search for green plant trimers. *Photochem. Photobiol. Sci.* **4**, 1091–1094 (2005).
- Boekema, E. J. et al. Evidence for a trimeric organization of the photosystem I complex from the thermophilic cyanobacterium *Synechococcus* sp. *FEBS Lett.* **217**, 283–286 (1987).
- Hippler, M., Drepper, F., Farah, J. & Roach, J. D. Fast electron transfer from cytochrome c6 and plastocyanin to photosystem I of *Chlamydomonas reinhardtii* requires *PsaF*. *Biochemistry* **36**, 6343–6349 (1997).
- Nelson, N. & Yocum, C. Structure and function of photosystems I and II. *Annu. Rev. Plant Biol.* **57**, 521–565 (2006).
- Merchant, S. & Sawaya, M. R. The light reactions: a guide to recent acquisitions for the picture gallery. *Plant Cell* **17**, 648–663 (2005).
- Kerfeld, C. A. & Krogmann, D. W. Photosynthetic cytochromes c in cyanobacteria, algae and plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 397–425 (1998).
- Rumeau, D., Peltier, G. & Cournac, L. Chlororespiration and cyclic electron flow around PSI during photosynthesis and plant stress response. *Plant Cell Environ.* **30**, 1041–1051 (2007).
- Bailey, S. et al. Alternative photosynthetic electron flow to oxygen in marine *Synechococcus*. *Biochim. Biophys. Acta* **1777**, 269–276 (2008).
- Partensky, F., La Roche, J., Wyman, K. & Falkowski, P. G. The divinyl-chlorophyll *a/b*-protein complexes of two strains of the oxyphototrophic marine prokaryote *Prochlorococcus*—characterization and response to changes in growth irradiance. *Photosynth. Res.* **51**, 209–222 (1997).
- Dammeyer, T., Bagby, S. C., Sullivan, M. B., Chisholm, S. W. & Frankenberg-Dinkel, N. Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. *Curr. Biol.* **18**, 442–448 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank M. Rosenberg for technical support, D. Rusch, J. Zehr and S. Bench for sharing genomic data, D. Lindell and R. Sorek for encouragement and discussions, and U. Pick for the comments on cyclic photosynthesis. This work was supported in part by grants 1203/06 (O.B.) and 356/06 (N.N.) from the Israel Science Foundation, by the Henry Taub Award for Academic Excellence, and by the Technion V.P.R. Fund-Henri Gutwirth Promotion of Research Fund (O.B.).

Author Contributions A.A. devised the initial idea for the project. I.S. and O.B. conceived the experiments. I.S. wrote the code and analysed the raw data, and together with F.G., R.Y.P., E.V.K., Y.I.W., N.N. and O.B. performed the bioinformatics. F.R. collected DNA and phage concentrates from the Northern Line Islands. A.A., N.A.-I. and M.H. conducted the molecular biology experiments, I.S., F.P., E.V.K., N.N. and O.B. co-wrote the paper.

Author Information The PSI sequences reported here have been deposited with GenBank under accession numbers EU926752–EU926761 (overlapping PCR) and GQ268816 (long PCR). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to O.B. (beja@tx.technion.ac.il).

Changes of mind in decision-making

Arbora Resulaj^{1,2}, Roozbeh Kiani³, Daniel M. Wolpert¹ & Michael N. Shadlen³

A decision is a commitment to a proposition or plan of action based on evidence and the expected costs and benefits associated with the outcome. Progress in a variety of fields has led to a quantitative understanding of the mechanisms that evaluate evidence and reach a decision^{1–3}. Several formalisms propose that a representation of noisy evidence is evaluated against a criterion to produce a decision^{4–8}. Without additional evidence, however, these formalisms fail to explain why a decision-maker would change their mind. Here we extend a model, developed to account for both the timing and the accuracy of the initial decision⁹, to explain subsequent changes of mind. Subjects made decisions about a noisy visual stimulus, which they indicated by moving a handle. Although they received no additional information after initiating their movement, their hand trajectories betrayed a change of mind in some trials. We propose that noisy evidence is accumulated over time until it reaches a criterion level, or bound, which determines the initial decision, and that the brain exploits information that is in the processing pipeline when the initial decision is made to subsequently either reverse or reaffirm the initial decision. The model explains both the frequency of changes of mind as well as their dependence on both task difficulty and whether the initial decision was accurate or erroneous. The theoretical and experimental findings advance the understanding of decision-making to the highly flexible and cognitive acts of vacillation and self-correction.

Decision-making spans a vast range of types and complexity, from choosing your partner or deciding whether to dive left or right to save a goal to simply deciding when to lift your finger. Studies of simple perceptual decisions have provided insight into the neurobiological mechanisms responsible for decision-making in both monkeys and humans (for reviews, see refs 1–3, 10). These studies often require a binary choice between two possible stimulus categories, such as leftward or rightward motion. Psychophysical and neural data¹ support models, termed drift–diffusion⁶, random walk^{5,7} and race⁸, in which a decision is made when the accumulated noisy evidence (decision variable) reaches a criterion level, termed a decision bound. Such an accumulation process explains both the accuracy of decisions over a range of difficulty levels as well as the time required to make the decisions⁹. These models are naturally viewed as an extension of signal detection theory and Bayesian inference to streams of data over time^{4,11}. One important limitation of the models is that they fail to explain why a decision-maker might change their mind after an initial decision has been taken. In some instances, such changes can lead to the correction of an initial error^{12,13}. Here we develop a task in which we can monitor changes of mind. We then extend the bounded-diffusion framework to explain both the frequency and the pattern of changes of mind.

Three naïve participants observed a moving random-dot stimulus and made decisions about the direction of motion (leftward or rightward), which they indicated by moving a handle to either a leftward or rightward target (Fig. 1a). Critically, the moving dots were extinguished as soon as the subjects initiated their movement

(Fig. 1b) and, hence, subjects could not acquire new evidence during their movement. The choice at initiation (initial hand trajectory) and reaction times as a function of task difficulty (coherence of dot motion) were explained by a model of bounded drift–diffusion (Fig. 2, black curves) consistent with previous studies in humans and monkeys^{1,9,14}. According to this model, evidence is accumulated until it reaches one of two bounds (corresponding to leftward and rightward decisions), which determines the choice and decision time.

Although no further visual information was available after movement initiation, the hand trajectories (Fig. 1c) gave a clear indication that in some trials observers changed their minds. That is, subjects generated a curved hand path that initially was on course to reach one target, but changed direction during the movement to finish at the other target. Although some changes of mind resulted in errors, the majority corrected an initial error. Changes of mind reliably improved accuracy (Fig. 2, top row: black and red circles correspond to the initial and final choices, respectively) for all three subjects by improving sensitivity to motion ($P < 0.006$ for each subject).

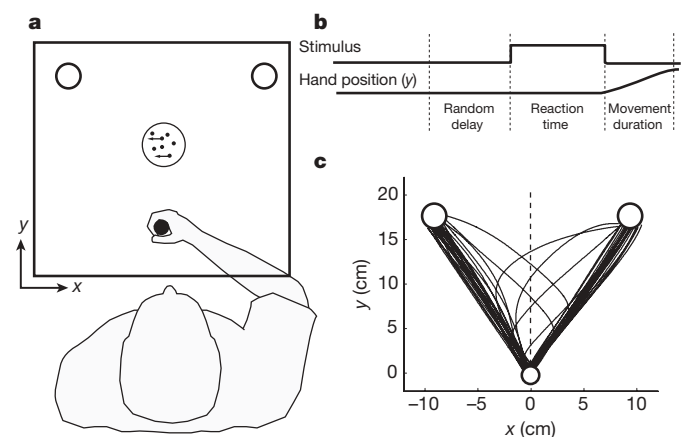


Figure 1 | Experimental set-up. **a**, Schematic of the visual display (rectangle). Subjects held the handle of a robotic interface (filled circle, shown here in the ‘home’ position) and moved to either a leftward or a rightward circular target depending on the perceived motion direction of a central random-dot display. A mirror system prevented subjects from seeing their arm. **b**, The time course of events that make up a trial. Each trial started when the subject’s hand was in the home position. After a random delay, the dots became visible and the subject could view the moving dot stimulus for as long as they needed (up to 2 s). Subjects indicated the direction of dot motion by moving to the leftward or rightward target. As soon as the subjects moved from the home position, the motion stimulus vanished. The trial ended when the subject reached one of the two targets. **c**, Sample hand trajectories from one subject. Most trajectories extend directly from the home position (bottom circle) to one of the choice targets. In a fraction of trials, the trajectories change course during the movement, indicating a change of mind.

¹Computational and Biological Learning Laboratory, Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK. ²Howard Hughes Medical Institute, Janelia Farm Research Campus, 19700 Helix Drive, Ashburn, Virginia 20147, USA. ³Howard Hughes Medical Institute, National Primate Research Center and Department of Physiology and Biophysics, University of Washington, Seattle, Washington 98195, USA.

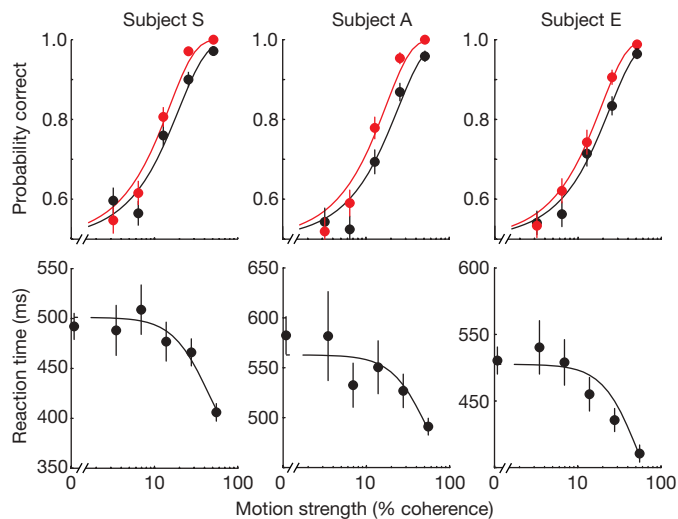


Figure 2 | Accuracy improves through changes of mind. Data are from three subjects (S, A and E). The top row shows that the probability of a correct decision at initiation (black) is lower than at termination (red) for almost all motion strengths. The bottom row shows that reaction times are longer for weaker motion strengths. Solid curves are fits to the data of the bounded-accumulation model (fraction of variance explained by the model fit, R^2 , for subjects S, A and E are respectively 0.96, 0.95 and 0.98 for initial decision, 0.98, 0.96 and 0.99 for final decision, and 0.92, 0.74 and 0.87 for reaction time). In this model, processing after initial commitment leads to an improvement in performance during the post-initiation phase. Error bars, s.e.m.

The observation is seemingly paradoxical. If there is information available to make a better decision, it might be expected to influence the initial decision. Every normative, 'ideal-observer'-based theory of decision-making would posit the decision as an inference made on the available evidence. The paradox is resolved if the decision-maker does not use all of the available evidence to make the initial choice but can tap into further information in the period between commitment to the initial response and termination of the movement.

Although the stimulus vanishes upon movement initiation, there is information in the processing pipeline that is potentially available to the decision-maker after movement initiation. Sensory- and motor-processing latencies ensure that not all of the information available from stimulus onset to movement initiation contributes to the decision. The sum of these latencies, termed the non-decision time (t_{nd}), was estimated to be 300–400 ms in our experiments (Supplementary Table 1 and Methods). Single-unit recordings from the lateral intraparietal area of the macaque in eye-movement versions of this task suggest that the non-decision time includes sensory and motor delays of around 220 ms and 80 ms, respectively^{15,16}. We proposed that the unused information could be processed after the brain has committed to an initial choice, thereby requiring an extension of the bounded-diffusion mechanism that includes post-initiation processing.

An analysis of the motion evidence leading to the subjects' choices supports this hypothesis. Each stimulus was a noisy sequence of random dots, which led to rapid fluctuations in the motion evidence, as quantified by 'motion energy'^{16,17} favouring left or right. For each trial, we removed the average motion energy associated with that motion strength and direction, leaving only the moment-to-moment fluctuations about the mean. We then averaged those residuals to look for evidence in the stimulus in support of the subjects' initial choice. The stimulus fluctuations immediately after stimulus onset supported the initial choice (Fig. 3a, left-hand blue curve: average over first 150 ms is positive; $P < 0.0001$), whereas the fluctuations in the final few hundred milliseconds had little bearing on the choice. For each subject, we identified the time point at which the average

came within 1 s.e. of zero (arrows), thus providing an empirical estimate of non-decision time. The motion-energy filtering induces a delay of 50–150 ms (Fig. 3a, inset). Taking this into account, the initial choices depend on the earliest information in the stimulus, but ignore an epoch on the order of t_{nd} .

The pattern was different for the subset of trials in which there is a change of mind. The early information from the stimulus provided weaker support for the initial choice (Fig. 3a, left-hand red trace) and exhibited a negative trend near the time of initiation (Fig. 3a, right-hand red trace), in support of the final, changed decision. The motion energy in this later epoch was significantly more negative relative to that in the remaining trials ($P < 0.0001$). The observation that motion energy supports both the initial and final choices provides evidence against two main alternatives to post-initiation processing: (1) change of decision based on recall and/or reconsideration of evidence acquired before initiation¹⁸, and (2) correction of an initial motor error perhaps due to confusion about the stimulus–response mapping¹². The analysis instead supports a non-decision time in which information from the stimulus arrives too late to affect an initial decision but is present to refine it after the brain has committed to a particular response and action.

We next considered how this extended processing could explain the pattern of changes of mind in the data. In particular, we wished to explain the proportion of changes to correct and to erroneous choices as a function of motion strength (Fig. 3b, red and black symbols, respectively). A seemingly optimal solution to the problem is to suppose the subject wishes to use changes of mind to maximize the percentage of correct final choices. Then the subject ought to continue to accumulate evidence about direction until there is no more to be had (that is, until time t_{nd}) and to decide in favour of the more likely direction. This formulation holds regardless of the trade-off between speed and accuracy underlying the initial choice. This idea fails to explain our findings: it predicts too many changes and it would defer them to the end of the evidence stream, which is clearly not the case (for example, early changes of mind, Fig. 1c). Because the subject must complete a hand movement, the optimal solution is likely to incorporate motor costs (energy) associated with larger corrections nearer the end of the movement. This idea can be realized by incorporating new bounds in the post-decision period to change or reaffirm an initial decision based on some criterion, thereby allowing changes to occur earlier in the movement. We considered a variety of models (Methods). The most parsimonious of these is illustrated in Fig. 3c. In this model, once the initial bound has been reached and a decision made, evidence continues to accumulate until it either reaches a new 'change-of-mind' bound or a time deadline terminates post-initiation processing. The decision rule is to change only if the accumulated evidence reaches the change-of-mind bound and to reaffirm otherwise. The offsets of the new bound and the deadline (two parameters) were fitted to account for the changes of mind as a function of coherence (Fig. 3b, curves).

For all three subjects, the model fits imply that upon termination of the initial decision, the subjects set a new bound at a level that would necessitate a reversal of the sign of the accumulated evidence. The amount of evidence required for a subject to change their mind (B_{Δ} , Supplementary Table 1) differed by $\sim 30\%$ across subjects, which explains the variation in the pattern of their changes. In all cases, the existence of this change-of-mind bound led to a significant improvement in the fits, in comparison with using all the available information (that is, no bound and choice based on the sign of the decision variable after t_{nd} ; $P < 0.003$ for all subjects, likelihood-ratio test). The deadline produced by the fit suggests that subjects avail themselves of most of the information in the processing pipeline. The model captures the complex dependence of post-initiation changes on both the motion strength and the initial decision ($R^2 = 0.63$ – 0.85 and 0.76 – 0.99 for changes to correct and incorrect choices, respectively). Changes of mind were most frequent at intermediate motion

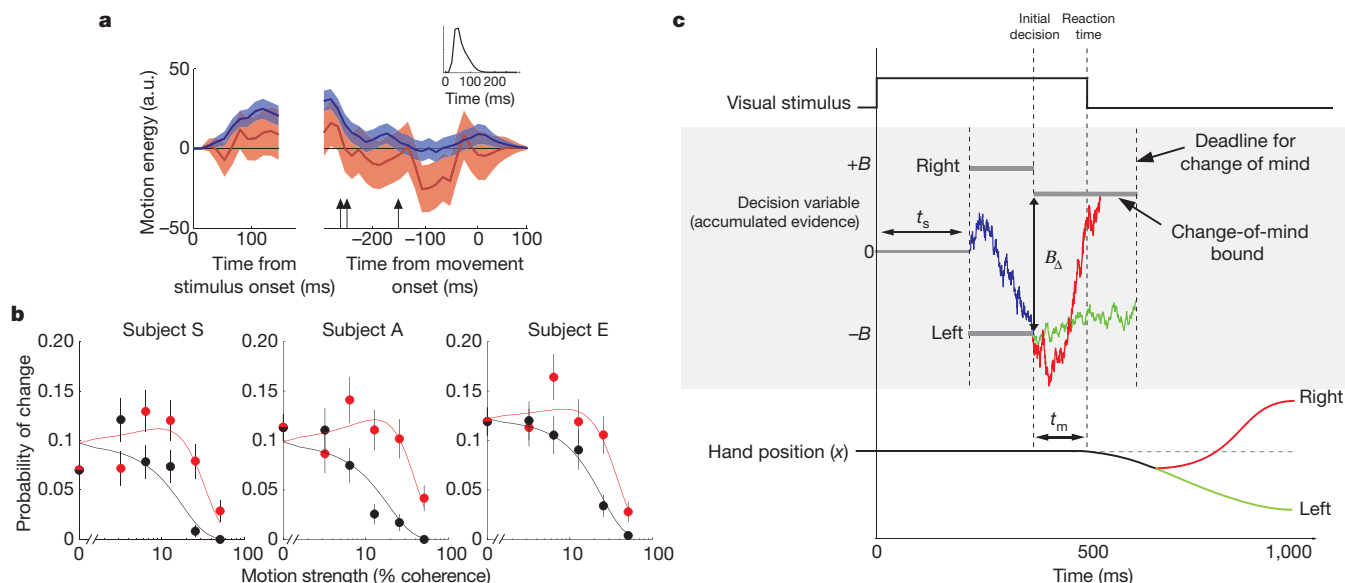


Figure 3 | A bounded-accumulation model of decision-making with post-initiation processing explains changes of mind. **a**, Influence of motion-energy fluctuations on initial and final decisions. Data are shown for all the trials (blue) and the subset of trials with a change of mind (red) aligned at stimulus onset (left) and movement onset (right). Motion-energy fluctuations were obtained by applying a filter to the sequence of random dots shown in each trial and subtracting the mean for all trials sharing the same motion strength and direction (Methods). The residual fluctuations are designated positive if they support the direction of the initial decision. Shading indicates s.e.m. Arrows indicate the time preceding movement initiation at which the average motion-energy fluctuations for each subject falls to within 1 s.e. of zero. Inset, impulse response for the filter used to calculate motion energy. a.u., arbitrary units. **b**, The model explains the probability of changes of mind from incorrect to correct choices (model, red curves; data, red symbols) and changes of mind from correct to incorrect choices (model, black curves; data, black symbols) as a function of stimulus

strengths when the initial choice was erroneous. The model offers an intuitive explanation for this. Viewed as a decision process beginning at the initial decision bound, there is a higher probability of reaffirming the initial choice, because the accumulated evidence is far from the change-of-mind bound. A change of mind therefore requires strong evidence in the short time available for post-initiation processing to move the accumulated evidence to the change-of-mind bound. Such strong evidence ought to arrive when the initial choice is an error and when the motion is strong. However, if the motion is very strong, initial errors are rare.

Our central finding is that the same data stream may be sampled at different moments to support different decisions and, hence, a change of mind. As a further test of this idea, we placed the timing of the initial decision under experimental control. This allowed us to isolate changes of mind from the strategies governing the trade-off of speed and accuracy of initial decisions in the reaction-time experiment. Instead of responding when ready, subjects were trained to time the initiation of their movement so that it coincided with an expected auditory beep. The stimulus motion began at a random time 200–2,000 ms (mean, 440 ms) before the beep and ended at the beep or at movement initiation, whichever occurred first (Methods). This experiment therefore tested whether our suggested framework generalizes to a situation in which the time of the initial choice is determined by an exogenous cue. The results of this experiment, which are summarized in Supplementary Figs 1–3, confirm the finding that subjects base their initial choice on early evidence but can avail themselves of additional evidence in the processing pipeline to revise this choice. These data also conform to a variant of the bounded-accumulation mechanism with post-initiation processing (Methods and Supplementary Figs 2 and 3).

coherence. Error bars, s.e.m. **c**, Information flow diagram showing visual stimulus and neural events leading to a decision and a possible change of mind. The example illustrates a rightward motion stimulus that gives rise to an initial incorrect leftward choice with reaction time around 500 ms. The visual stimulus gives rise to a decision variable (blue trace) that is the accumulation of noisy evidence. This governs the initial choice and decision time. The initial decision is complete when a 'Right' or 'Left' bound is crossed (that is, $\pm B$ of evidence has accumulated). Data from neural recordings^{15,16} suggest that the delay from motion onset to the beginning of the accumulation (t_s) is around 200 ms, and the delay from the initial decision to movement initiation (t_m) is around 80 ms. The time of the termination is around the mean decision time for the three subjects. Further accumulation takes place on the evidence still in the processing pipeline; if the accumulated evidence reaches the opposite change-of-mind bound then the decision is reversed (red), and if the deadline is reached then the decision is confirmed (green).

We expect the change-of-mind mechanism to apply under a wide variety of conditions if there is time pressure to respond. When two of our subjects were instructed to perform the reaction-time experiment more slowly, their initial decisions were more accurate and there were fewer changes of mind (data not shown). The pattern was explained by the same model with higher initiation bounds⁹. Also, because in our study the subject must complete an arm movement, the optimal solution is likely to trade off accuracy against motor costs (energy) associated with larger corrections nearer the end of the movement. Determining the optimal bounds for such a trade-off will require the coupling of concepts derived from theories of optimal feedback control¹⁹ and decision-making models. We suspect that more complex situations, for example in which movements must be timed more precisely or when a correction is more costly, might necessitate both a reaffirmation bound and bounds whose heights vary over time.

Our proposed mechanism cannot explain all changes of mind. For example, it cannot explain corrections of initial errors that arise from confusion about stimulus–response associations¹². Furthermore, a change that depends on retrieval of information from memory or incorporation of a new decision policy (for example values) would require elaboration of the model. Presumably these types of vacillations could be based on more complex processes that involve memory retrieval or application of a new criterion on a stored decision variable.

Advances in understanding the neurobiology of decision-making have benefited from simple perceptual tasks^{18,20,21}, but the same principles appear to underlie decisions related to foraging², gambling²², social selection²³ and probabilistic reasoning²⁴. The common principle is that the representation of information bearing on choice is imperfect, thus inviting the application of some criterion against

which to judge the evidence. The class of bounded-diffusion models^{5–7,25,26} extends this theory of signal classification⁴ to data streams and thus incorporates time costs as well^{27,28}. An unexpected virtue of such models demonstrated by our experiment is that a part of the data stream that is not used to make the decision can nonetheless support revision after a response is initiated.

This formalism provides a view of decision-making in which subjects can exploit the expectation that late-arriving information may or may not be useful to refine a decision or action. We suspect that when a change of decision is costly, energetically or otherwise, subjects will naturally tend to shun this strategy and opt for longer initial decision times. A change is precluded when an action is ballistic, for instance when a subject makes an eye movement to a choice target^{9,15}. In these instances, a change of mind can only lead to a post-decision regret²⁹ or possibly a learning signal even in the absence of overt feedback. On the other hand, a variety of complex motor sequences might benefit from early initiation premised on the expectation of additional information that is in the pipeline. It is well known that the initiation and final specifications of a movement can be dissociated in time³⁰. What we have shown here is that when these processes act on the same data stream, they can lead to a change in a decision. We speculate that a common neural mechanism explains refinement of a movement after initiation and what we experience cognitively as a change of mind about a proposition.

METHODS SUMMARY

Three naive subjects performed the main experiment. The local ethics committee approved the protocol. Subjects moved a handle in the horizontal plane. A mirror overlaid virtual images from a computer monitor onto the plane of the movement. The hand position was displayed as a small blue circle. After a random delay, a dynamic random-dot stimulus appeared (Fig. 1). In each trial, the direction of motion was randomly chosen to be leftward or rightward. Task difficulty was varied randomly by controlling the fraction of coherently moving dots. The subjects were instructed to judge the net direction of motion as quickly and as accurately as they could, and to move the handle to either a leftward or rightward target. The motion stimulus was extinguished when the movement was initiated. The trial ended when the subject reached one of the targets. Subjects performed an initial training session of at least 500 trials followed by 1,500 test trials.

We recorded the hand trajectories at 1,000 Hz. For each trial, we measured the reaction time and the final target selection. Normally hand movements for easy trials (high coherence) were straight to the target. A change of mind was reflected in a trajectory that initially travelled towards one target but ended at the other. We calculated the area between the hand path and the line from the starting position to the midpoint between the two targets. A change of mind was detected if the area swept out by the hand on the side opposite the final chosen target exceeded 0.1 cm². This criterion was based on a control experiment using 100% coherent motion. We were therefore able to determine for each trial the choice at both initiation and termination of the movement.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 April; accepted 10 July 2009.

Published online 19 August 2009.

- Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
- Sugrue, L. P., Corrado, G. S. & Newsome, W. T. Choosing the greater of two goods: neural currencies for valuation and decision making. *Nature Rev. Neurosci.* **6**, 363–375 (2005).
- Schall, J. D. Neural basis of deciding, choosing and acting. *Nature Rev. Neurosci.* **2**, 33–42 (2001).
- Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (Wiley, 1966).

- Laming, D. R. J. *Information Theory of Choice-Reaction Times* (Wiley, 1968).
- Ratcliff, R. & Rouder, J. N. Modelling response times for two-choice decisions. *Psychol. Sci.* **9**, 347–356 (1998).
- Link, S. W. The relative judgment theory of two choice response time. *J. Math. Psychol.* **12**, 114–135 (1975).
- Smith, P. L. & Vickers, D. The accumulator model of two-choice discrimination. *J. Math. Psychol.* **32**, 135–168 (1988).
- Palmer, J., Huk, A. C. & Shadlen, M. N. The effect of stimulus strength on the speed and accuracy of a perceptual decision. *J. Vis.* **5**, 376–404 (2005).
- Heekeren, H. R., Marrett, S. & Ungerleider, L. G. The neural systems that mediate human perceptual decision making. *Nature Rev. Neurosci.* **9**, 467–479 (2008).
- Beck, J. M. et al. Probabilistic population codes for Bayesian decision making. *Neuron* **60**, 1142–1152 (2008).
- Rabbitt, P. & Vyas, S. Processing a display even after you make a response to it. How perceptual errors can be corrected. *Q. J. Exp. Psychol. A* **33**, 223–239 (1981).
- Rabbitt, P. M. Error correction time without external error signals. *Nature* **212**, 438 (1966).
- Smith, P. L. & Ratcliff, R. Psychology and neurobiology of simple decisions. *Trends Neurosci.* **27**, 161–168 (2004).
- Roitman, J. D. & Shadlen, M. N. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.* **22**, 9475–9489 (2002).
- Kiani, R., Hanks, T. D. & Shadlen, M. N. Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *J. Neurosci.* **28**, 3017–3029 (2008).
- Adelson, E. H. & Bergen, J. R. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* **2**, 284–299 (1985).
- Romo, R., Hernandez, A., Zainos, A., Lemus, L. & Brody, C. D. Neuronal correlates of decision-making in secondary somatosensory cortex. *Nature Neurosci.* **5**, 1217–1225 (2002).
- Todorov, E. Optimality principles in sensorimotor control. *Nature Neurosci.* **7**, 907–915 (2004).
- Parker, A. J. & Newsome, W. T. Sense and the single neuron: probing the physiology of perception. *Annu. Rev. Neurosci.* **21**, 227–277 (1998).
- Uchida, N., Kepecs, A. & Mainen, Z. F. Seeing at a glance, smelling in a whiff: rapid forms of perceptual decision making. *Nature Rev. Neurosci.* **7**, 485–491 (2006).
- Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876–879 (2006).
- Deaner, R. O., Khera, A. V. & Platt, M. L. Monkeys pay per view: adaptive valuation of social images by rhesus macaques. *Curr. Biol.* **15**, 543–548 (2005).
- Yang, T. & Shadlen, M. N. Probabilistic reasoning by neurons. *Nature* **447**, 1075–1080 (2007).
- Usher, M. & McClelland, J. L. The time course of perceptual choice: the leaky, competing accumulator model. *Psychol. Rev.* **108**, 550–592 (2001).
- Wong, K. F. & Wang, X. J. A recurrent network mechanism of time integration in perceptual decisions. *J. Neurosci.* **26**, 1314–1328 (2006).
- Gold, J. I. & Shadlen, M. N. Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron* **36**, 299–308 (2002).
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P. & Cohen, J. D. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* **113**, 700–765 (2006).
- Stuphorn, V., Taylor, T. L. & Schall, J. D. Performance monitoring by the supplementary eye field. *Nature* **408**, 857–860 (2000).
- Ghez, C., Hening, W. & Favilla, M. Gradual specification of response amplitude in human tracking performance. *Brain Behav. Evol.* **33**, 69–74 (1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Wellcome Trust, the European grant SENSOPAC IST-2005-028056, Howard Hughes Medical Institute and US National Eye Institute grant EY11378. We thank A. Faisal, H. Vincent, I. Howard and J. Ingram for their assistance. M.N.S. thanks Trinity College, Cambridge, for support.

Author Contributions D.M.W. and M.N.S. planned the experiments. A.R. performed the experiments. All authors analysed and interpreted results, and all authors wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.N.S. (shadlen@uw.edu).

METHODS

Behavioural task. Four naive subjects (three male and one female) provided informed consent and participated in the experiment. The local ethics committee approved the protocol. Three subjects performed each of the reaction-time and cued-movement experiments (two subjects, S and E, performed both, with the reaction-time experiment first). Subjects were seated and used their preferred hand to hold the handle of a vBOT manipulandum³¹ that was free to move in the horizontal plane (Fig. 1a). Subjects were prevented from seeing their arm by a mirror that was used to overlay virtual images of a video display (updated at 75 Hz) onto the plane of the movement. A chin- and headrest ensured a viewing distance of 40 cm. The hand position was displayed as a small blue circle (radius, 0.5 cm).

The time course of a trial in the reaction-time experiment is shown in Fig. 1b. A trial began when the subject's hand was in the home position (circle of radius 1 cm; Fig. 1a). After a random delay, sampled from a truncated exponential distribution (range, 0.7–1.0 s; mean, 0.82 s), a dynamic random-dot stimulus appeared at the centre of the screen within a circular aperture subtending 5° of visual angle. The motion stimulus is described in detail in previous studies¹⁵. In each trial, the direction of motion was randomly chosen to be leftward or rightward. The stimulus density was 15.6 dots deg⁻² s⁻¹. Dots were displayed for one video frame and then either replaced at a random position or displaced to the left or right three video frames (40 ms) later. This displacement would produce a speed of 7.1° s⁻¹. Thus the positions of the dots in frame four, say, were correlated only with the displaced dots in frames one and/or seven but with none of the dots in frames two, three, five and six. The probability that each dot would be displaced as opposed to randomly replaced, termed the per cent coherence, determined the task difficulty and was selected randomly from the set (0%, 3.2%, 6.4%, 12.8%, 25.6%, 51.2%).

The subjects were instructed to judge the direction of the moving random dots as quickly and as accurately as they could, and to reach to a corresponding circular target (one on the left and one on the right; radius, 1.5 cm; 20 cm from the starting position and 28° from the midline; Fig. 1a). Critically, when the movement was initiated—that is, the hand crossed the boundary of the home-position circle—the random-dot stimulus was extinguished. Subjects were required to reach the target with a movement duration of 500 ± 200 ms. The trial ended when the subject reached one of the targets. Subjects were provided with visual feedback of whether they had made the correct choice (for the 0% coherence trials, half of the trials were randomly designated 'correct'). Subjects were instructed to maintain fixation throughout at a small cross in the centre of the dot aperture—the targets were large enough that they could be easily reached using peripheral vision. Subjects performed an initial training session of at least 500 trials followed by 1,500 test trials.

In the cued-movement task, subjects heard five beeps equally spaced in time (500-ms spacing) and were required to initiate movement on the fourth beep and reach the target on the fifth beep (Supplementary Fig. 1a). Random-dot motion began at a random interval before the fourth beep (truncated exponential distribution: range, 0.2–2 s; mean, 0.44 s). The motion display was extinguished on the fourth beep or at the time of movement initiation if the subject slightly anticipated the beep. Feedback was provided to maintain movement initiation and termination within ±100 ms of the fourth and fifth beeps, respectively. Again, subjects were given feedback of whether they had made the correct choice. Subjects performed an initial training session of 500 trials followed by 2,000 test trials.

Data analysis. We recorded the hand trajectories at 1,000 Hz. For each trial, we quantified the reaction time (time to movement initiation from start of motion stimulus) and the final target selection. In addition, we developed a measure, based on the hand trajectories, of whether subjects had changed their decision during the movement. Normally hand movements for easy trials (high coherence) were straight to the target (Fig. 1c). A change of mind was reflected in a trajectory that initially travelled towards one target but ended at the other. We calculated the area between the hand path and the line from the starting position to the bisector of the two targets. A change of mind was deemed to have occurred if the area swept out by the hand on the side opposite the final chosen target exceeded 0.1 cm² and the point of maximum horizontal deviation was outside the home position. This criterion was chosen on the basis of a control experiment with two of our subjects using the reaction-time condition but with 100%-coherent motion stimuli. We expected to see few, if any, changes of mind under this condition and in fact observed two change-of-mind trials out of 400, both of which were obvious lapses with swept areas at least three times larger than the criterion, suggesting that our method of determining changes of mind is conservative. We were therefore able to determine for each trial the choice at both initiation and termination of the movement.

Modelling. For the reaction-time experiment (Figs 2 and 3), we adapted a bounded-accumulation model (Fig. 3c) to explain the initial- and final-choice

frequencies (Fig. 3b). We first explain the model for the initial choices and then expand it to explain changes of mind.

For the initial choices, the model posits that evidence accumulates from a starting point, y_0 , until it reaches an upper or lower bound ($\pm B$), which determines the initial choice and decision time. The increments of evidence are idealized as normally distributed random variables with unit variance per second and mean $\mu = kC + \mu_0$, where C is signed motion strength (a positive value corresponding to rightward motion and negative value corresponding to leftward motion); k , B , y_0 and μ_0 are free parameters. The parameters B and k explain the trade-off between the speed and the accuracy of the initial choices; μ_0 and y_0 are respectively drift and starting-point offsets, which explain bias for one of the choices. The bias terms were not necessary for all subjects (Supplementary Table 1).

This formulation leads to the following simplification³², which may help to provide an intuition for the effect of motion strength on initial choice and reaction time. If $y_0 = 0$, the probability of a rightward initial choice is

$$P_{\text{right}} = [1 + \exp(-2\mu B)]^{-1}$$

and the mean decision time is

$$t_d = \frac{B}{\mu} \tanh(\mu B)$$

The reaction time incorporates additional latencies from stimulus onset to the beginning of the bounded-accumulation process and from the termination of the process to the beginning of the motor response. The sum of these latencies, the non-decision time t_{nd} , is an additional parameter of the model such that the measured reaction time is $t_d + t_{nd}$, which we set for each direction choice.

Because the stimulus duration in each trial equals the reaction time, there is additional evidence from the stimulus that is potentially available for processing after the brain has committed to an initial choice. The model incorporates this additional information as follows. When the initial decision ends, the accumulation continues (from $\pm B$) until either a second, post-initiation change-of-mind bound is crossed, in which case the decision is reversed, or a temporal deadline is exceeded, in which case the initial decision is reaffirmed (Fig. 3c). The height of this new bound was offset by B_d from the initiation bound. A value of $B_d = B$ would imply that a change of mind occurs when the evidence changes sign, and a value of $B_d = 2B$ would imply that a change requires an amount of net evidence represented by the initial bounds. The values for our subjects were between B and $2B$.

The fits to the initial choices and reaction times provide the sensitivity parameter (k), initial bounds (B) and non-decision times (t_{nd}) used in the post-initiation analyses. We then considered a series of plausible models for the post-initiation phase. These models were intended to explain the observed initial and final choices (bivariate observations: left–left, left–right and so on) given fixed values for k , B and t_{nd} . The strategy ensures that all comparison models are on equal footing and that the number of parameters for post-initiation is small. We compared an 'optimal' model using all available evidence (no additional degrees of freedom (d.f.)), a single flat change-of-mind bound (d.f. = 1), a flat change-of-mind bound with a deadline (as described above; d.f. = 2), flat bounds for change of mind and for reaffirmation (d.f. = 2), and variants of these models with quadratic collapsing bounds (an extra 1–2 d.f. to parameterize the collapse). We used a likelihood-ratio test for nested models and supported these comparisons using the Bayes information criterion³³. On the basis of these comparisons, we adopted the simplest model that accounted for all the subjects' data (Fig. 3c): one with a single change-of-mind bound and a cut-off that would censor late information acquired during t_{nd} . The parameters for this model are shown in the final two rows of Supplementary Table 1. All fits were performed using maximum-likelihood methods. Model choice probabilities and reaction-time distributions were derived from numerical solutions of Fokker–Planck equations for the bounded-diffusion process³⁴.

Although it appears that a large number of parameters were used to model the initial and final choices, the strategy is conservative and intuitive. We used six parameters for the fits to the initial choices and reaction times to ensure that the estimates of parameters that affect the post-initiation phase (k , B and t_{nd}) were as accurate as possible. A model with just three parameters gives acceptable fits for the initial choices and reaction times for all three subjects, but the additional parameters explain the small biases in two of the subjects and the 4–10-ms difference in t_{nd} for leftward and rightward choices. Although several of these terms have negligible effects for one or more subjects (Supplementary Table 1), they produce more accurate estimates of k , B and t_{nd} . As noted above, the simple two-parameter model used to fit the post-initiation data was supported by an extensive model comparison. To perform this model comparison with as much power and sensitivity as possible, it was necessary to place all models on equal footing by supplying the best possible values for the inherited parameters (k , B

and t_{nd}). In particular, we did not want to justify a more complicated model (for example one with collapsing bounds) simply because the additional degrees of freedom could explain residual error in k and B . Our strategy is conservative in that it tends to reduce the explanatory power of more complex models for changes of mind.

We also performed a cross-validation analysis to ensure that the large number of parameters in our fit to the reaction-time data did not lead to overfitting. We split each subject's data set into two equal halves (random permutation of trials at each motion strength) and fitted each separately. We used the fits from one half to predict the other half of the data. The cross-validation fits, goodness of fit and parameter estimates are shown in Supplementary Fig. 4 and Supplementary Tables 3 and 4. The similarity of the predictions and fits provides reassurance that the model is not overparameterized.

A simpler version of the model was used to fit data from the cued-movement experiment (Supplementary Figs 2 and 3). Here the non-decision time, t_{nd} , delimits the portion of the data stream available for the initial choice. In a trial in which the stimulus is displayed for a time t_{stim} , subjects can use $t_d = t_{stim} - t_{nd}$ of the data stream (or no information if the stimulus duration is shorter than the non-decision time) to determine their initial choice, and a further t_{nd} (or the stimulus duration if shorter than t_{nd}) to potentially revise their decision. Put simply, the initial choice is governed by the sign of the decision variable after t_d of diffusion, whether or not it has terminated. Post-initiation processing occurs on the remaining data stream until either the left or right choice bound is reached. The same symmetric bounds were used before and after initiation. A key difference from the reaction-time experiment is that once the accumulated evidence has reached a bound, the diffusion process terminates and there is no opportunity for a change of mind. Thus, only non-terminated decisions after time t_d are eligible for a change of mind. This seems sensible because, unlike in the reaction-time experiment, the subject does not choose the time of initiation. Termination of the process is tantamount to accepting that the level of evidence is sufficient for a choice. Model fits (for k , B and t_{nd} , see Supplementary Table 2) were obtained using maximum-likelihood methods. Because initiation was timed to coincide with an external beep in this experiment, the main effect of the bounds was to curtail the improvement in accuracy that would be expected for perfect integration for long times t_d (ref. 16). The initial- and final-choice probabilities were derived by numerical solution of Fokker–Planck equations for each trial, using the same stimulus durations as in the data set.

Statistical analysis. Unless otherwise stated, P values are based on t statistics constructed from parameter estimates and their associated standard errors. We calculated the standard errors by using the inverse Hessian from maximum-likelihood fits wherever possible, or a bootstrap procedure³⁵ when the numerical solution of the Fokker–Planck equation did not support accurate calculation of the Hessian. For the fraction of t_{nd} , we report the 95% confidence interval (method of fiducial limits³⁶, likelihood-ratio test) because this parameter is bounded by zero and one. The R^2 values accompanying the model fits were calculated as one minus the fraction of unexplained variance for the data points displayed in the graphs. To evaluate the differences between initial- and final-choice probabilities, we did not rely on the model in Fig. 3 but instead performed logistic regression. Accordingly, the probability of choosing right is given by

$$P_{\text{right}} = [1 + \exp(-b_0 - b_1 C - b_2 I - b_3 IC)]^{-1}$$

where C is the signed motion strength, I is an indicator variable (zero for initial choice and one for the final choice) and b_i are fitted coefficients. To test for improved sensitivity (accuracy) with changes of mind, we evaluated the null hypothesis $\{H_0: b_3 \leq 0\}$. An alternative formulation—probability correct as a function of unsigned motion strength—confirmed the statistical significance of this analysis as well as the analysis of the cued-motion experiment.

For the motion-energy analyses, we extracted a time series from the sequence of random dots shown in each trial by applying a filter for rightward and leftward motion with passband centred at 1.0 cyc deg⁻¹ and 7.1 Hz, thus matching the speed and dot displacement in our stimulus (for details, see refs 16, 17). The difference in these time series represents momentary evidence in favour of one or the other choice. To combine data across trials, we removed the average motion energy associated with each trial's motion strength and direction. We then applied a sign convention so that positive fluctuations are in the direction of the subject's initial choice. The graphs in Fig. 3a and Supplementary Fig. 2b show these averaged residuals, time-locked to either stimulus onset or movement initiation.

For the statistical analysis of the motion energy time-locked to movement initiation, we used the data from all trials (blue curves) to identify the point in time (for each subject) at which stimulus motion fluctuations no longer influence the initial choice, using an arbitrary value of 1 s.e. from zero. This procedure gives a model-free estimate of t_{nd} . We analysed the motion energy from the change-of-mind trials from this time until movement initiation. To test whether the total motion energy in an epoch differed significantly from zero, we applied a permutation test (randomization of the sign of motion energy in each trial)³⁵. To compare the motion energy in change-of-mind and reaffirmation trials, we applied a bootstrap procedure. We calculated the total motion energy in the change-of-mind trials using the epoch defined above and compared this with the distribution of values obtained in randomly resampled trials without change of mind over the identical epochs. This bootstrap comparison compensated for a lack of power due to there being relatively few change-of-mind trials (for example, neither of the trends in the left-hand red curves of Fig. 3a and Supplementary Fig. 2b are significantly different from zero).

31. Howard, I. S., Ingram, J. N. & Wolpert, D. M. A modular planar robotic manipulandum with end-point torque control. *J. Neurosci. Methods* **181**, 199–211 (2009).
32. Shadlen, M., Hanks, T., Churchland, A., Kiani, R. & Yang, T. in *Bayesian Brain: Probabilistic Approaches to Neural Coding* (ed. Doya, K. et al.) 209–237 (MIT Press, 2006).
33. Kass, R. E. & Wasserman, L. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *J. Am. Stat. Assoc.* **90**, 928–934 (1995).
34. Risken, H. *The Fokker–Planck Equation: Methods of Solution and Applications* 2nd edn (Springer, 1989).
35. Efron, B. *The Jackknife, the Bootstrap and Other Resampling Plans* (Society for Industrial and Applied Mathematics, 1982).
36. Wang, Y. H. Fiducial intervals: what are they? *Am. Stat.* **54**, 105–111 (2000).

The avian Z-linked gene *DMRT1* is required for male sex determination in the chicken

Craig A. Smith¹, Kelly N. Roeszler¹, Thomas Ohnesorg¹, David M. Cummins², Peter G. Farlie¹, Timothy J. Doran² & Andrew H. Sinclair¹

Sex in birds is chromosomally based, as in mammals, but the sex chromosomes are different and the mechanism of avian sex determination has been a long-standing mystery^{1–3}. In the chicken and all other birds, the homogametic sex is male (ZZ) and the heterogametic sex is female (ZW). Two hypotheses have been proposed for the mechanism of avian sex determination. The W (female) chromosome may carry a dominant-acting ovary determinant^{4–6}. Alternatively, the dosage of a Z-linked gene may mediate sex determination, two doses being required for male development (ZZ)^{7,8}. A strong candidate avian sex-determinant under the dosage hypothesis is the conserved Z-linked gene, *DMRT1* (doublesex and mab-3-related transcription factor 1)^{9–11}. Here we used RNA interference (RNAi) to knock down *DMRT1* in early chicken embryos. Reduction of *DMRT1* protein expression *in ovo* leads to feminization of the embryonic gonads in genetically male (ZZ) embryos. Affected males show partial sex reversal, characterized by feminization of the gonads. The feminized left gonad shows female-like histology, disorganized testis cords and a decline in the testicular marker, SOX9. The ovarian marker, aromatase, is ectopically activated. The feminized right gonad shows a more variable loss of *DMRT1* and ectopic aromatase activation, suggesting differential sensitivity to *DMRT1* between left and right gonads. Germ cells also show a female pattern of distribution in the feminized male gonads. These results indicate that *DMRT1* is required for testis determination in the chicken. Our data support the Z dosage hypothesis for avian sex determination.

Two different RNAi approaches were used to knock down endogenous *DMRT1* transcripts, delivered into living chicken embryos via the avian retroviral vector RCASBP(B)¹². The virus carried the green fluorescent protein (GFP) gene to monitor viral spread, with a *DMRT1* microRNA (miRNA) in the 3' untranslated region (UTR) of the transgene (designated miRNA563), or with an internal U6 RNA polymerase promoter driving expression of a different short hairpin RNA (shRNA) independently of GFP (designated shRNA343)¹³. The first construct, miRNA563, targeted exon three of *DMRT1*, whereas the second construct, shRNA343, targeted the DNA-binding domain of exon two. The miRNA and shRNA constructs delivered similar levels of robust GFP expression and knockdown of exogenous *DMRT1* protein in cultured chicken DF1 cells (Fig. 1). Cells infected with RCASBP(A) strain virus expressing only the *DMRT1* complementary DNA showed strong *DMRT1* overexpression (Fig. 1a). Cells pre-infected with virus carrying a non-silencing scrambled miRNA control followed by RCASBP(A)*DMRT1* still showed robust *DMRT1* protein expression (Fig. 1b). In contrast, cells pre-infected with *DMRT1* microRNA (miRNA563) followed by *DMRT1* showed knockdown of the protein (Fig. 1c), and an 80% reduction in *DMRT1* transcript compared to cells treated with scrambled control

miRNA (Fig. 1d). Similarly, DF1 cells co-transfected with plasmids expressing shRNA343 and *DMRT1*–GFP fusion protein showed a 70% reduction in GFP reporter expression compared to controls

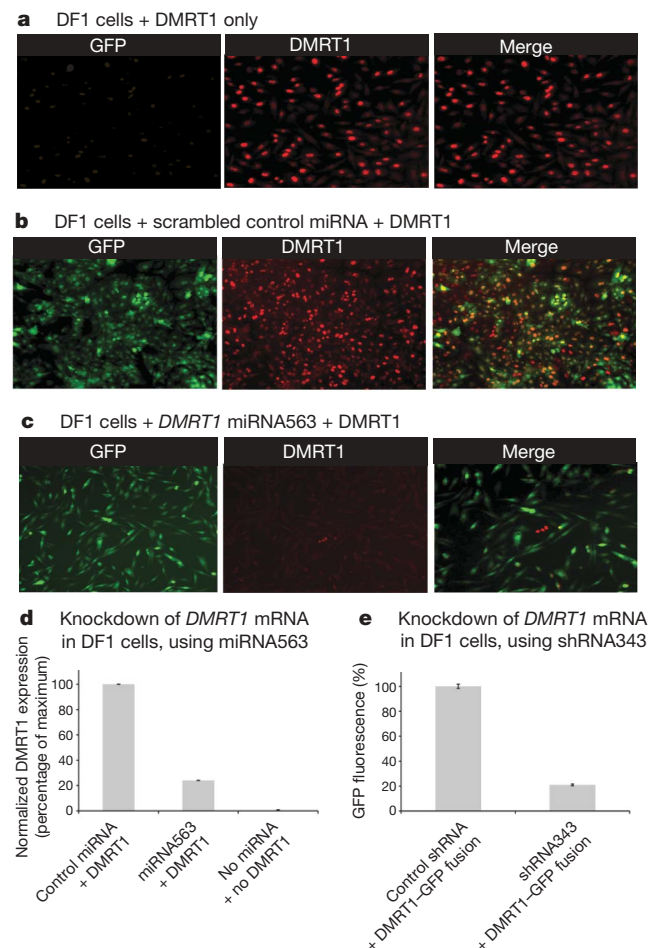


Figure 1 | Knockdown of *DMRT1* expression *in vitro*, using RCASBP(B) virus to deliver miRNA or shRNA against *DMRT1*. **a**, DF1 cells infected with *DMRT1* only, showing no GFP expression but robust expression of *DMRT1*. **b**, Cells infected with *DMRT1* and a non-silencing control scrambled miRNA with GFP reporter, showing widespread GFP and *DMRT1* protein expression. **c**, Cells infected with *DMRT1* and *DMRT1* miRNA563, showing GFP expression and *DMRT1* knockdown. **d**, Knockdown of *DMRT1* mRNA with *DMRT1* plus miRNA563, compared to controls. Mean \pm s.d. **e**, Knockdown of *DMRT1*–GFP fusion protein with *DMRT1*–GFP and shRNA343 plasmids. Mean \pm s.e.m.

¹Murdoch Children's Research Institute and Department of Paediatrics, The University of Melbourne, Royal Children's Hospital, Melbourne, Victoria 3052, Australia. ²CSIRO Livestock Industries, Australian Animal Health Laboratory, Geelong, Victoria 3220, Australia.

(Fig. 1e). RNase protection assays confirmed expression of mature *DMRT1* knockdown short interfering RNAs (siRNAs) in DF1 cells infected with RCASBP(B) expressing shRNA343 (Supplementary Fig. 1).

In the chicken embryo, the gonads form on the mesonephric kidneys around day 3.5 of incubation. Sexual differentiation into testes or ovaries begins at day 6 and is normally advanced by day 10. Embryos infected with virus at day 0 showed global GFP reporter expression by day 10, including widespread expression in the urogenital system and in sectioned gonads (Supplementary Fig. 2). RNase protection assays of these day-10 gonads confirmed expression of the mature *DMRT1* knockdown siRNAs (Supplementary Fig. 1). We infected 550 embryos with *DMRT1* knockdown viruses (Table 1). Of these embryos, 24% showed GFP fluorescence on macroscopic examination of the gonads. Gonads were categorized into groups on the basis of their overall GFP expression at the macroscopic level: low, medium or high. Quantitative PCR with reverse transcription (RT–PCR) analysis revealed an inverse correlation between gonadal GFP and endogenous *DMRT1* gene expression. Gonads with high levels of GFP, and hence miRNA or shRNA delivery, showed significantly reduced *DMRT1* messenger RNA, whereas those with lower GFP gene expression showed more modest *DMRT1* mRNA reduction (Supplementary Fig. 3). Male embryos with low or no GFP expression appeared normal (confirmed by histology or immunostaining) and were excluded from further analysis. In contrast, genetic males with high GFP expression showed feminized gonads, as assessed by gonadal histology, immunofluorescence and quantitative RT–PCR ($n = 27$) (Table 1).

Feminization of genetically male chicken embryos with *DMRT1* knockdown viruses is shown schematically and by histology in Fig. 2. Gonadal development in embryos treated with non-silencing scrambled control miRNA and showing high GFP expression was normal ($n = 22$). Control and *DMRT1* knockdown females showed typical asymmetric development that was characterized by a large left ovary and smaller regressing right gonad (Fig. 2a). At the histological level, the left ovary had a well developed outer cortex, populated with germ cells, and a vacuolated medulla riddled with lacunae (cavities) (Fig. 2b, c). Scrambled control males had bilateral testes (Fig. 2d, e). Within both testes, well-developed seminiferous cords occupied the medulla and a thin surface epithelium was present (Fig. 2f). In contrast, *DMRT1* knockdown males with high GFP expression showed varying degrees of female-like asymmetry at the macroscopic level (Fig. 2g, h). The smaller right gonad had either seminiferous cords or poorly organized cords (Fig. 2i). The left gonad was strongly feminized, with a vacuolated medulla and thickened outer cortex (Fig. 2j), as in control females.

Gonads from embryos with high levels of GFP (and hence *DMRT1* knockdown) were assessed for *DMRT1* and marker gene expression. In control embryos, *DMRT1* protein was uniformly expressed in the nuclei of developing Sertoli and germ cells within testis cords (Fig. 3a). In contrast, male embryos (ZZ) treated with two different *DMRT1* knockdown constructs showed variably reduced *DMRT1* protein expression, disrupted testis cord formation and ectopic female gene expression. The extent of *DMRT1* knockdown and testis cord disruption varied among embryos, but was more pronounced in

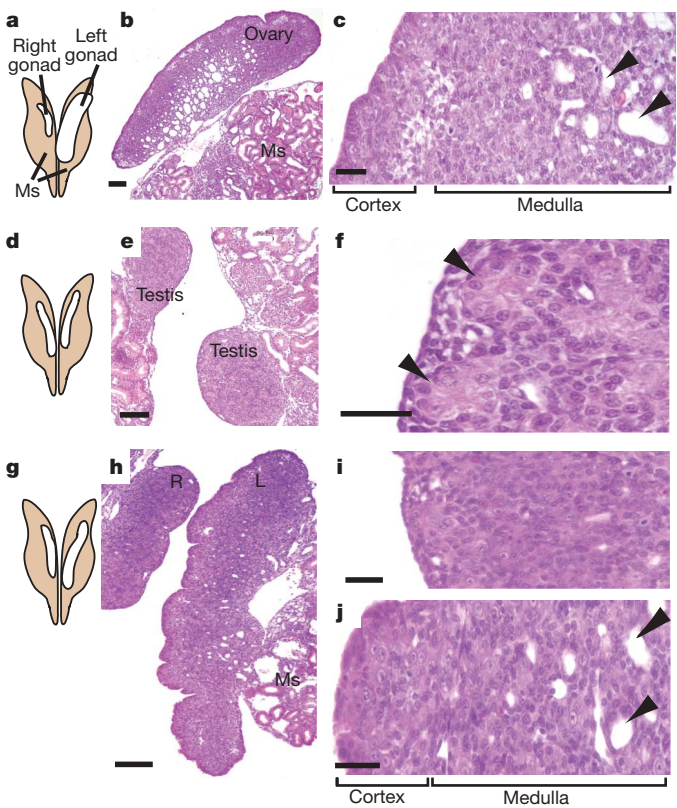


Figure 2 | Feminization of male gonads following knockdown of *DMRT1*. **a**, Semi-schematic view of a control female urogenital system. **b**, Section of a control left ovary. Ms, mesonephric kidneys. **c**, High-power view of a control left ovary, showing vacuolated medulla with lacunae (for example, arrowheads) and thickened cortex. **d**, Schematic view of a control male urogenital system. **e**, Section through a control male urogenital system, showing paired testes. **f**, High-power view of a control testis, showing well-developed seminiferous cords (arrowheads). **g**, Schematic view of a male urogenital system following *DMRT1* knockdown. **h**, Section through a male knockdown urogenital system, showing large female-like left gonad and smaller male-like right gonad. **i**, High-power view of a right gonad from a *DMRT1* knockdown male, showing poorly organized cords. **j**, High-power view of the left gonad of a *DMRT1* knockdown male, showing female-like organization. The medulla is vacuolated, with lacunae (arrowheads), whereas the cortex is thickened. **b**, **e**, **h**, Scale bars, 100 μ m; **c**, **f**, **i**, **j**, scale bars, 25 μ m.

the presence of shRNA343 compared to miRNA563. Gonadal *DMRT1* protein expression was either greatly reduced throughout both left and right gonads (Fig. 3b), or its expression was irregular, in embryos treated with shRNA343. Some expression was still present in germ cells, which seem to silence the viral vector¹⁴. In male embryos treated with miRNA563, gonads generally showed partial feminization, characterized by an average or small-sized right testis with normal *DMRT1* expression, and a larger feminized left gonad with reduced *DMRT1* expression (Fig. 3c). Quantitative RT–PCR analysis confirmed that *DMRT1* mRNA expression was reduced by more than

Table 1 | Day-10 chicken embryos used for analysis of *DMRT1* knockdown

Viral treatment at the blastoderm stage	Viable day-10 embryos (both sexes)	Number of embryos (both sexes) showing high GFP expression in gonads	Total number of feminized genetic males (ZZ)/males with high GFP
RCASBP(B)	60% ($n = 65$)	40	0/22
Scrambled control			
RCASBP(B)	45% ($n = 350$)	40	15/17
<i>DMRT1</i> miRNA563			
RCASBP(B)	65% ($n = 200$)	28	12/15
<i>DMRT1</i> shRNA343			

GFP expression was assessed macroscopically by examining whole urogenital systems under a dissecting microscope equipped with fluorescence optics. GFP fluorescence was classified as low, medium or high. Feminization was assessed by gonadal histology, immunofluorescent marker expression or quantitative RT–PCR.

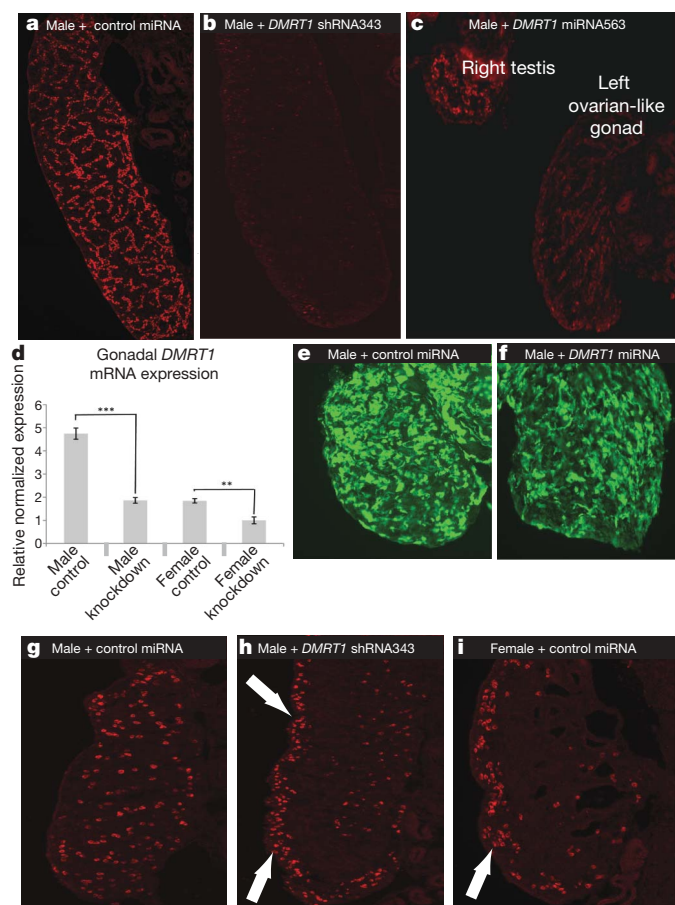


Figure 3 | Feminization of male chicken embryos following knockdown of *DMRT1*. **a**, *DMRT1* protein expression in a control male testis. **b**, Reduction of *DMRT1* protein expression in the left gonad of a feminized male following *DMRT1* knockdown. **c**, Partial sex reversal in a male following *DMRT1* knockdown, showing normal *DMRT1* expression in a small right testis and greatly reduced *DMRT1* expression in a larger left ovarian-like gonad. **d**, *DMRT1* mRNA expression in control and knockdown gonads (mean \pm s.e.m.; *** $P < 0.001$; ** $P < 0.01$; $n \geq 3$). **e**, GFP reporter expression in a control male gonad treated with scrambled miRNA. **f**, GFP reporter expression in a knockdown male gonad treated with miRNA563. **g**, Chicken Vasa homologue (CVH) staining, showing distribution of germ cells within the interior (testis cords) of a control male gonad, treated with scrambled miRNA. **h**, Female-like cortical distribution of germ cells (arrows) in a male gonad treated with *DMRT1* knockdown miRNA563 and immunostained for CVH. **i**, Cortical germ cell distribution in the left ovary of a control female (arrow), immunostained for CVH.

60% in male embryos treated with miRNA563 or shRNA343 (Fig. 3d). Both scrambled control and knockdown gonads showed robust GFP reporter expression in tissue sections (Fig. 3e, f).

In male gonads with significantly reduced *DMRT1* expression, germ cells showed a cortical (female-like) distribution pattern, compared to a medullary cord distribution in control males (Fig. 3g–i). In control females, ovarian development was normal. *DMRT1* was expressed at low levels in both gonads, with the exception of higher expression in cortical germ cells of the left ovary. In genetic females treated with *DMRT1* miRNA and shRNA, endogenous *DMRT1* expression was lower (Fig. 3d), but the gonads nevertheless appeared normal, with typical asymmetry (Supplementary Fig. 4). This suggests that *DMRT1* is not essential for chicken ovarian development, consistent with the human and mouse data¹⁵.

Gonads were further examined for the expression of male and female markers. A key gene involved in testicular differentiation is *SOX9*, which is upregulated in all male vertebrate embryos that have been examined so far, including birds¹⁶. In day-10 control male embryos, *SOX9* protein was expressed normally in the nuclei of

Sertoli cells. Female gonads (either control or *DMRT1* knockdown) lacked *SOX9* expression (Fig. 4a, b). In genetic males treated with *DMRT1* miRNA563, *SOX9* protein expression was variably reduced, reflecting disrupted testis cords (Fig. 4c). In male embryos treated with shRNA343, reduction of *SOX9* expression was more marked (Fig. 4d, e). At the mRNA level, *SOX9* expression was significantly reduced in *DMRT1* knockdown gonads relative to controls (Fig. 4f). *DMRT1* may therefore have a role in the activation or maintenance of *SOX9* expression during testis determination in the chicken embryo (a role partially filled by the *SRY* gene in mammals).

Genetic male chicken embryos treated with *DMRT1* miRNA also showed ectopic activation of the female marker, aromatase. Aromatase enzyme is normally expressed only in female gonads, where it synthesizes the oestrogen that is required for ovarian differentiation in birds¹⁷. Aromatase is never detected in normal male embryonic gonads. In control and *DMRT1* knockdown female embryos, aromatase was strongly expressed in the medulla of both left and right gonads (Fig. 5a). No expression was seen in male controls treated with scrambled miRNA (Fig. 5b). However, aromatase was ectopically activated in *DMRT1* knockdown males. In males treated with shRNA343, both the left and right gonads showed ectopic aromatase expression (Fig. 5c). However, only the left gonad expressed aromatase in males treated with miRNA563. This difference may be attributed to the more robust *DMRT1* knockdown seen with the former construct. In some individuals with partial feminization, areas of reduced or absent *DMRT1* expression correlated with ectopic aromatase and female-like lacunae (Fig. 5d, e). Ectopic activation of aromatase in male gonads treated with *DMRT1* miRNA or shRNA was confirmed by quantitative RT–PCR (Fig. 5f). These findings indicate that increased *DMRT1* expression in male gonads normally suppresses aromatase and hence female development. This effect could be direct, or indirect via repression of the forkhead transcription factor, *FOXL2*, which is postulated to regulate aromatase¹⁸. An indirect effect via *FOXL2* is supported by quantitative RT–PCR, which showed ectopic

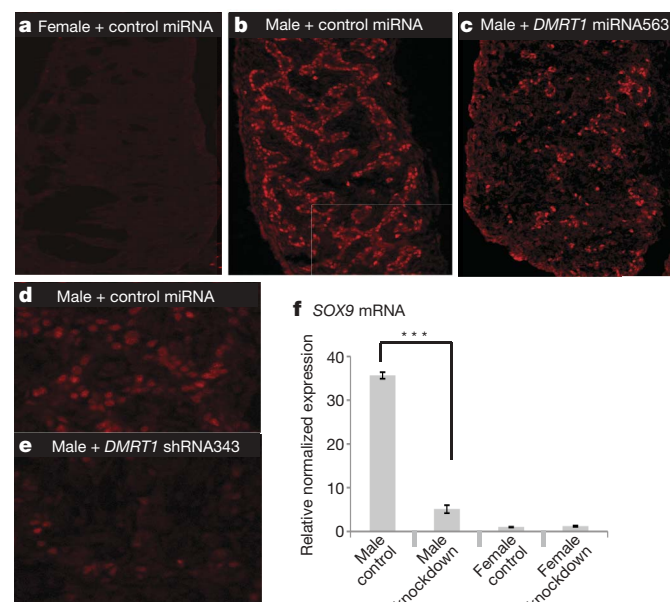


Figure 4 | Downregulation of *SOX9* in male gonads following *DMRT1* knockdown. **a**, No *SOX9* protein expression in a control female. **b**, Normal *SOX9* protein expression in organized testis cords of a control male. **c**, Reduced *SOX9* expression and disorganized testis cords in a male treated with *DMRT1* miRNA563, longitudinal section. **d**, High-power view of *SOX9*-positive Sertoli cells in a control male gonad. **e**, High-power view of a male gonad treated with shRNA343, showing downregulation of *SOX9* expression. **f**, Downregulation of *SOX9* mRNA in *DMRT1* knockdown gonads compared to controls. (Quantitative RT–PCR. Mean \pm s.e.m.; *** $P < 0.001$; $n \geq 3$).

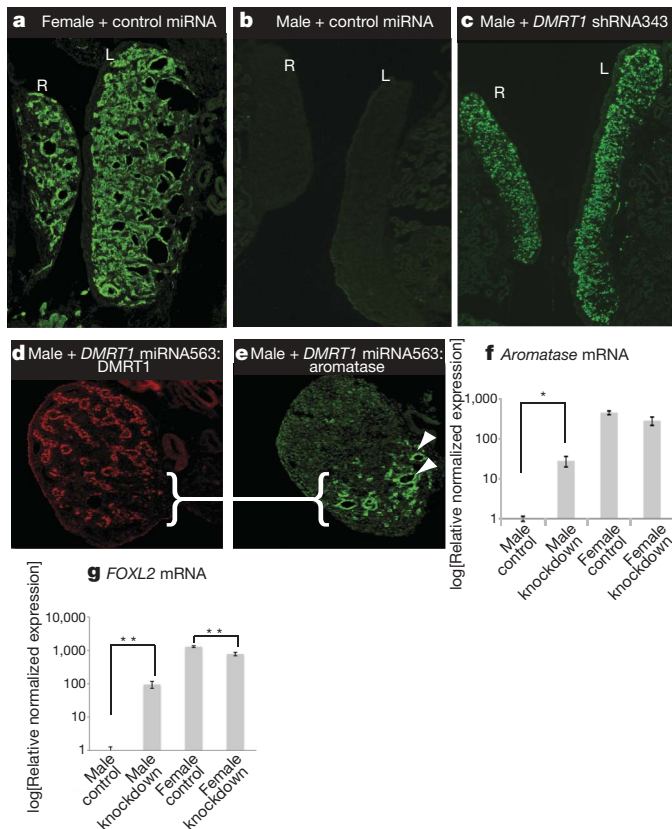


Figure 5 | Ectopic expression of female markers in male gonads following *DMRT1* knockdown. **a**, Strong bilateral expression of aromatase protein in control female gonads. **b**, No aromatase expression in control male gonads. **c**, Ectopic aromatase expression in male gonads following treatment with *DMRT1* shRNA343. L, left gonad; R, right gonad. **d**, Partial *DMRT1* expression in a male embryo treated with miRNA563, showing normal *DMRT1* protein at one pole and reduced *DMRT1* at the other pole (bracketed). **e**, The region with reduced *DMRT1* in **d** shows ectopic aromatase expression, and female-like lacunae (arrows). **f**, Ectopic expression of *Aromatase* (also known as *CYP19A1*) mRNA in male knockdown gonads relative to controls (mean \pm s.e.m.; * $P < 0.05$; $n \geq 3$). **g**, Activation of *FOXL2* mRNA expression in *DMRT1* knockdown gonads relative to controls (mean \pm s.e.m.; * $P < 0.01$; $n \geq 3$).

expression of *FOXL2* mRNA in genetic males treated with *DMRT1* miRNA (Fig. 5g).

These results indicate that *DMRT1* has a key role in chicken testis determination. Treatment of genetic male chicken embryos with two different knockdown constructs, targeting two different *DMRT1* exons, results in feminization of the gonads by day 10 of development. The viral construct shRNA343 resulted in more pronounced feminization than miRNA563. This may be because RCASBP(B)shRNA343 uses an internal U6 promoter, compared to the viral long terminal repeat in RCASBP(B)miRNA563. In addition, the two viral constructs target different *DMRT1* exons, with shRNA343 targeting exon two and shRNA563 targeting exon three. Recent data suggest that the chicken *DMRT1* gene may be alternatively spliced¹⁹, and shRNA343 was designed to target all potential isoforms. Thus, shRNA343 may be a more potent knockdown construct. Nevertheless, when miRNA563 was used, the feminizing effect was more apparent in the left versus the right gonad. There may be intrinsic differences between the left and right gonad with respect to *DMRT1* sensitivity, with the left being more susceptible. The female-like size asymmetry frequently seen in male knockdown gonads suggests that *DMRT1* may antagonize the lateralizing effects of *PITX2*, a transcription factor recently shown to be responsible for the asymmetric development of embryonic chicken gonads^{20,21}.

It is predicted from results reported here that the converse experiment, overexpression of *DMRT1*, will masculinize genetically female (ZW) embryos. However, *DMRT1* overexpression experiments using the retroviral system described here caused embryo lethality by day 4 (before gonadal sex differentiation), presumably owing to its global effects. This problem could potentially be addressed by using a tissue-specific promoter to direct *DMRT1* overexpression in the gonads.

Although Z-linked *DMRT1* is required for testis development, it is possible that another Z-linked gene lies upstream of this gene in the avian male-determining pathway. However, this seems unlikely. It is also possible that a female determinant lies on the avian W sex chromosome, which has few true genes. The best W-linked candidate sex determinant, *HINTW* (also known as *ASW* or *WPKCI*), does not induce female development when overexpressed in male embryos¹⁴. However, an alternative W-linked ovary-determining gene may exist.

Our results support the Z dosage hypothesis for avian sex determination²². Under this hypothesis, a higher dosage of *DMRT1* initiates testicular differentiation in male embryos, activating *SOX9* expression and suppressing aromatase. *DMRT1* fulfils the requirements expected of an avian master sex-determining gene. It is sex-linked and conserved on the Z sex chromosome of all birds examined, including the basal ratites (emus, ostriches and so on)²³. It is expressed exclusively in the urogenital system before gonadal sex differentiation in chicken embryos, with higher expression in males²⁴, and knockdown leads to feminization. In other vertebrates, *DMRT1* is also implicated in testis development. *DMRT1*-null mutant mice have impaired postnatal testis development¹⁵, and deletions of *DMRT1* in humans cause testicular dysgenesis²⁵. In reptiles with temperature sex determination, *DMRT1* expression is upregulated during the thermo-sensitive period when sex is being determined, and only at male-producing temperatures^{26,27}. In the medaka fish, *Oryzias latipes*, a duplicated copy of *DMRT1*, *dmy/dmrt1b*, is the master testis determinant²⁸, and a W-linked copy, *dmw*, is involved in ovarian development in an amphibian, *Xenopus laevis*²⁹. Our data provide evidence that *DMRT1* is the male sex determinant in birds, confirming a pervasive role for DM (*Doublesex/Mab-3*) domain genes in vertebrate sex determination.

METHODS SUMMARY

Preparation of RCASBP knockdown constructs. The avian retroviral vector RCASBP(B) was used to deliver knockdown sequences directed specifically against chicken *DMRT1* mRNA. See Supplementary Methods for further details. *DMRT1* knockdown viruses were propagated in chicken DF1 cells, collected and titred as described previously¹⁴.

Infection of chicken embryos. Chicken (*Gallus gallus domesticus*) embryos of a susceptible strain were infected with concentrated viruses as previously reported¹⁴.

Cell culture and RNase protection assays. Chicken DF1 cells were used to propagate and test RCASBP(B) viruses carrying *DMRT1* knockdown constructs¹⁴. An RNase protection assay was performed to detect *DMRT1* shRNA343 expression using a specific RNA probe. See Supplementary Methods for further details.

Immunofluorescence. Tissues were briefly fixed in 4% paraformaldehyde-PBS, cryosectioned and processed for immunofluorescence as described previously^{14,30}.

Quantitative RT-PCR. RNA was extracted from gonads, reverse-transcribed and real-time PCR was performed using the Universal Probe Library (UPL) system and Light Cycler 480 probe master mix (Roche). All samples were normalized against *HPRT* using the comparative C_t method ($\Delta\Delta C_t$), detailed in the Supplementary Methods.

Received 26 January; accepted 21 July 2009.

Published online 26 August 2009.

- Ellegren, H. Evolution of the avian sex chromosomes and their role in sex determination. *Trends Ecol. Evol.* **15**, 188–192 (2000).
- Mizuno, S. *et al.* Z and W chromosomes of chickens: studies on their gene functions in sex determination and sex differentiation. *Cytogenet. Genome Res.* **99**, 236–244 (2002).
- Smith, C. A., Roeszler, K. N., Hudson, Q. J. & Sinclair, A. H. Avian sex determination: what, when and where? *Cytogenet. Genome Res.* **117**, 165–173 (2007).

4. Smith, C. A. Avian sex determination: HINTS from the W sex chromosome? *Sex. Dev.* **1**, 278–285 (2007).
5. Arlt, D., Bensch, S., Hansson, B., Hasselquist, D. & Westerdaal, H. Observation of a ZZW female in a natural population: implications for avian sex determination. *Proc. R. Soc. Lond. B* **271** (Suppl. 4), 249–251 (2004).
6. Hori, T., Asakawa, S., Itoh, Y., Shimizu, N. & Mizuno, S. *Wpkci*, encoding an altered form of *PKCI*, is conserved widely on the avian W chromosome and expressed in early female embryos: implication of its role in female sex determination. *Mol. Biol. Cell* **11**, 3645–3660 (2000).
7. Nanda, I. *et al.* 300 million years of conserved synteny between chicken Z and human chromosome 9. *Nature Genet.* **21**, 258–259 (1999).
8. Smith, C. A. & Sinclair, A. H. Sex determination: insights from the chicken. *Bioessays* **26**, 120–132 (2004).
9. Nanda, I. *et al.* Conserved synteny between the chicken Z sex chromosome and human chromosome 9 includes the male regulatory gene *DMRT1*: a comparative (re)view on avian sex determination. *Cytogenet. Cell Genet.* **89**, 67–78 (2000).
10. Shan, Z. *et al.* Sex-specific expression of an evolutionarily conserved male regulatory gene, *DMRT1*, in birds. *Cytogenet. Cell Genet.* **89**, 252–257 (2000).
11. Smith, C. A., McClive, P. J., Western, P. S., Reed, K. J. & Sinclair, A. H. Conservation of a sex-determining gene. *Nature* **402**, 601–602 (1999).
12. Logan, M. & Tabin, C. Targeted gene misexpression in chick limb buds using avian replication-competent retroviruses. *Methods* **14**, 407–420 (1998).
13. Wise, T. G. *et al.* Characterization and comparison of chicken U6 promoters for the expression of short hairpin RNAs. *Anim. Biotechnol.* **18**, 153–162 (2007).
14. Smith, C. A., Roeszler, K. N. & Sinclair, A. H. Robust and ubiquitous GFP expression in a single generation of chicken embryos using the avian retroviral vector, RCASBP. *Differentiation* **77**, 473–482 (2009).
15. Raymond, C. S., Murphy, M. W., O'Sullivan, M. G., Bardwell, V. J. & Zarkower, D. *Dmrt1*, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. *Genes Dev.* **14**, 2587–2595 (2000).
16. Kent, J., Wheatley, S. C., Andrews, J. E., Sinclair, A. H. & Koopman, P. A. Male-specific role for *SOX9* in vertebrate sex determination. *Development* **122**, 2813–2822 (1996).
17. Elbrecht, A. & Smith, R. G. Aromatase enzyme activity and sex determination in chickens. *Science* **255**, 467–470 (1992).
18. Govoroun, M. S. *et al.* Isolation of chicken homolog of the *FOXL2* gene and comparison of its expression patterns with those of aromatase during ovarian development. *Dev. Dyn.* **231**, 859–870 (2004).
19. Zhao, Y., Lu, H., Yu, H., Cheng, H. & Zhou, R. Multiple alternative splicing in gonads of chicken *DMRT1*. *Dev. Genes Evol.* **217**, 119–126 (2007).
20. Ishimaru, Y. *et al.* Mechanism of asymmetric ovarian development in chick embryos. *Development* **135**, 677–685 (2008).
21. Guioli, S. & Lovell-Badge, R. *PITX2* controls asymmetric gonadal development in both sexes of the chick and can rescue the degeneration of the right ovary. *Development* **134**, 4199–4208 (2007).
22. Nanda, I., Schlegelmilch, K., Haaf, T., Scharl, M. & Schmid, M. Synteny conservation of the Z chromosome in 14 avian species (11 families) supports a role for Z dosage in avian sex determination. *Cytogenet. Genome Res.* **122**, 150–156 (2008).
23. Shetty, S., Kirby, P., Zarkower, D. & Graves, J. A. *DMRT1* in a ratite bird: evidence for a role in sex determination and discovery of a putative regulatory element. *Cytogenet. Genome Res.* **99**, 245–251 (2002).
24. Smith, C. A., Katz, M. & Sinclair, A. H. *DMRT1* is upregulated in the gonads during female-to-male sex reversal in ZW chicken embryos. *Biol. Reprod.* **68**, 560–570 (2003).
25. Ounap, K. *et al.* Three patients with 9p deletions including *DMRT1* and *DMRT2*: a girl with XY complement, bilateral ovotestes, and extreme growth retardation, and two XX females with normal pubertal development. *Am. J. Med. Genet. A* **130A**, 415–423 (2004).
26. Shoemaker, C., Ramsey, M., Queen, J. & Crews, D. Expression of *Sox9*, *Mis*, and *Dmrt1* in the gonad of a species with temperature-dependent sex determination. *Dev. Dyn.* **236**, 1055–1063 (2007).
27. Kettlewell, J. R., Raymond, C. S. & Zarkower, D. Temperature-dependent expression of turtle *Dmrt1* prior to sexual differentiation. *Genesis* **26**, 174–178 (2000).
28. Matsuda, M. *et al.* *DMY* is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* **417**, 559–563 (2002).
29. Yoshimoto, S. *et al.* A W-linked DM-domain gene, *DM-W*, participates in primary ovary development in *Xenopus laevis*. *Proc. Natl Acad. Sci. USA* **105**, 2469–2474 (2008).
30. Smith, C. A. *et al.* Cloning and expression of *R-Spondin1* in different vertebrates suggests a conserved role in ovarian development. *BMC Dev. Biol.* **8**, 72 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Tabin for providing the RCASBP retroviral vectors, and X.-J. Yang for advice on viral preparation. We also thank M. Clinton and H. Sang for methods of manipulating early chicken embryos, S. White for statistical advice and P. Farmer for the histology. P. Koopman provided comments on the manuscript. This work was supported by an Australian Research Council (ARC) DP grant held by C.A.S. and A.H.S., and an NH&MRC Program grant jointly held by A.H.S.

Author Contributions C.A.S. and A.H.S. initiated the project and supervised the study. C.A.S., K.N.R., T.O. and A.H.S. conceived and designed the experiments. P.G.F. designed the pRmiR vector, whereas D.M.C. and T.J.D. designed the RCASBP(B)GFP.U6.shRNA343 and associated shuttle vectors. C.A.S., K.N.R., T.O. and D.M.C. performed the experiments. C.A.S., K.N.R., T.O. and A.H.S. analysed the data and prepared the manuscript. All authors approved the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.A.S. (craig.smith@mcrl.edu.au).

LETTERS

Targeted capture and massively parallel sequencing of 12 human exomes

Sarah B. Ng¹, Emily H. Turner¹, Peggy D. Robertson¹, Steven D. Flygare¹, Abigail W. Bigham², Choli Lee¹, Tristan Shaffer¹, Michelle Wong¹, Arindam Bhattacharjee⁴, Evan E. Eichler^{1,3}, Michael Bamshad², Deborah A. Nickerson¹ & Jay Shendure¹

Genome-wide association studies suggest that common genetic variants explain only a modest fraction of heritable risk for common diseases, raising the question of whether rare variants account for a significant fraction of unexplained heritability^{1,2}. Although DNA sequencing costs have fallen markedly³, they remain far from what is necessary for rare and novel variants to be routinely identified at a genome-wide scale in large cohorts. We have therefore sought to develop second-generation methods for targeted sequencing of all protein-coding regions ('exomes'), to reduce costs while enriching for discovery of highly penetrant variants. Here we report on the targeted capture and massively parallel sequencing of the exomes of 12 humans. These include eight HapMap individuals representing three populations⁴, and four unrelated individuals with a rare dominantly inherited disorder, Freeman–Sheldon syndrome (FSS)⁵. We demonstrate the sensitive and specific identification of rare and common variants in over 300 megabases of coding sequence. Using FSS as a proof-of-concept, we show that candidate genes for Mendelian disorders can be identified by exome sequencing of a small number of unrelated, affected individuals. This strategy may be extendable to diseases with more complex genetics through larger sample sizes and appropriate weighting of non-synonymous variants by predicted functional impact.

Protein-coding regions constitute ~1% of the human genome or ~30 megabases (Mb), split across ~180,000 exons. A brute-force approach to exome sequencing with conventional technology⁶ is expensive relative to what may be possible with second-generation platforms³. However, the efficient isolation of this fragmentary genomic subset is technically challenging⁷. The enrichment of an exome by hybridization of shotgun libraries constructed from 140 µg of genomic DNA to seven microarrays was described previously⁸. To improve the practicality of hybridization capture, we developed a protocol to enrich for coding sequences at a genome-wide scale starting with 10 µg of DNA and using two microarrays. Our initial target was 27.9 Mb of coding sequence defined by CCDS (the NCBI Consensus Coding Sequence database)⁹. This curated set avoids the inclusion of spurious hypothetical genes that contaminate broader exome definitions¹⁰. The target is reduced to 26.6 Mb on exclusion of regions that are poorly mapped with our anticipated read length owing to paralogous sequences elsewhere in the genome (Supplementary Data 1).

We captured and sequenced the exomes of eight individuals previously characterized by the HapMap⁴ and Human Genome Structural Variation¹¹ projects. We also analysed four unrelated individuals affected with Freeman–Sheldon syndrome (FSS; Online Mendelian Inheritance in Man (OMIM) #193700), also called distal arthrogryposis type 2A, a rare autosomal dominant disorder caused by

mutations in *MYH3* (ref. 5). Unpaired, 76 base-pair (bp) reads¹² from post-enrichment shotgun libraries were aligned to the reference genome¹³. On average, 6.4 gigabases (Gb) of mappable sequence was generated per individual (20-fold less than whole genome sequencing with the same platform¹²), and 49% of reads mapped to targets (Supplementary Table 1). After removing duplicate reads that represent potential polymerase chain reaction artefacts¹⁴, the average fold-coverage of each exome was 51× (Supplementary Fig. 1). On average per exome, 99.7% of targeted bases were covered at least once, and 96.3% (25.6 Mb) were covered sufficiently for variant calling ($\geq 8\times$ coverage and Phred-like¹⁵ consensus quality ≥ 30). This corresponded to 78% of genes having $>95\%$ of their coding bases called (Supplementary Fig. 2 and Supplementary Data 2). The average pairwise correlation coefficient between individuals for gene-by-gene coverage was 0.87, consistent with systematic bias in coverage between individual exomes.

False positives and false negatives are critical issues in genomic resequencing. We assessed the quality of our exome data in four ways. First, comparing sequence-based calls for the eight HapMap exomes to array-based genotyping, we observed a high concordance with both homozygous (99.94%; $n = 219,077$) and heterozygous (99.57%; $n = 43,070$) genotypes (Table 1). Second, we compared our coding single-nucleotide polymorphism (cSNP) catalogue to ~1 Mb of coding sequence determined in each of the eight HapMap individuals by molecular inversion probe (MIP) capture and direct resequencing¹⁶. At coordinates called in both data sets, 99.9% of all cSNPs ($n = 4,620$) and 100% of novel cSNPs ($n = 334$) identified here were concordant, consistent with a low false discovery rate. Third, we compared the NA18507 cSNPs identified here to those called by recent whole genome sequencing of this individual¹², and found substantial overlap (Supplementary Fig. 3). The relative numbers of cSNPs called by only one approach, and the proportions of these represented in dbSNP, indicate that exome sequencing has equivalent sensitivity for cSNP detection compared to whole genome sequencing. Fourth, we compared our data to cSNPs in high-quality Sanger sequence of single haplotype regions from fosmid clones of the same HapMap individuals¹⁷. Most fosmid-defined cSNPs (38 of 40) were at coordinates with sufficient coverage in our data for variant calling. Of these, 38 of 38 were correctly identified as variant.

A comparison of our data to past reports on exonic¹⁸ or exomic⁸ array-based capture revealed roughly equivalent capture specificity, but greater completeness in terms of coverage and variant calling (Supplementary Table 2). These improvements probably arise from a combination of greater sequencing depth and differences in array designs and in experimental conditions for capture. Within the set of

¹Department of Genome Sciences, ²Department of Pediatrics, University of Washington, ³Howard Hughes Medical Institute, Seattle, Washington 98195, USA. ⁴Agilent Technologies, Santa Clara, California 95051, USA.

Table 1 | Sequence coverage and array-based validation

Individual	Covered $\geq 1\times$	Sequence called	Concordance with Illumina Human1M-Duo calls		
			Homozygous reference	Heterozygous	Homozygous non-reference
NA18507 (YRI)	26,477,161 (99.7%)	25,795,189 (97.1%)	23757/23762 (99.98%)	5553/5583 (99.46%)	3582/3592 (99.72%)
NA18517 (YRI)	26,476,761 (99.7%)	25,748,289 (97.0%)	23701/23705 (99.98%)	5575/5601 (99.54%)	3568/3579 (99.69%)
NA19129 (YRI)	26,491,035 (99.8%)	25,733,587 (96.9%)	23701/23708 (99.97%)	5482/5510 (99.49%)	3681/3690 (99.76%)
NA19240 (YRI)	26,486,481 (99.7%)	25,576,517 (96.3%)	23546/23551 (99.98%)	5600/5634 (99.40%)	3542/3549 (99.80%)
NA18555 (CHB)	26,475,665 (99.7%)	25,529,861 (96.1%)	23980/23984 (99.98%)	4877/4893 (99.67%)	3776/3786 (99.74%)
NA18956 (JPT)	26,454,942 (99.6%)	25,683,248 (96.7%)	24217/24221 (99.98%)	4890/4910 (99.59%)	3751/3760 (99.76%)
NA12156 (CEU)	26,476,155 (99.7%)	25,360,704 (95.5%)	23789/23794 (99.98%)	5493/5514 (99.62%)	3206/3213 (99.78%)
NA12878 (CEU)	26,439,953 (99.6%)	25,399,572 (95.6%)	23885/23891 (99.97%)	5413/5425 (99.78%)	3274/3292 (99.45%)
FSS10066 (Eur)	26,467,140 (99.7%)	25,546,738 (96.2%)	NA	NA	NA
FSS10208 (Eur)	26,461,768 (99.6%)	25,576,256 (96.3%)	NA	NA	NA
FSS22194 (Eur)	26,426,401 (99.5%)	25,454,551 (95.9%)	NA	NA	NA
FSS24895 (Eur)	26,478,775 (99.7%)	25,602,677 (96.4%)	NA	NA	NA

The number of coding bases covered at least $1\times$ and with sufficient coverage to variant call ($\geq 8\times$ and consensus quality ≥ 30) are listed for each exome, with the fraction of the aggregate target (26.6 Mb) that this represents in parentheses. For the eight HapMap individuals, concordance with array genotyping (Illumina Human1M-Duo) is listed for positions that are homozygous for the reference allele, heterozygous or homozygous for the non-reference allele (according to the array genotype). CEU, CEPH HapMap; CHB, Chinese HapMap; Eur, European-American ancestry (non-HapMap); JPT, Japanese HapMap; YRI, Yoruba HapMap. NA, Not applicable.

called positions, the high concordance with heterozygous array-based genotypes ($>99\%$) provides an estimate of our sensitivity for rare variant detection, as rare variants are overwhelmingly expected to be heterozygous. However, sensitivity was limited in that $\sim 4\%$ of known heterozygous genotypes were at coordinates where there was insufficient coverage to make a confident call.

There were 56,240 cSNPs called in one or more individuals, of which 13,347 were novel. On average, 17,272 cSNPs were called per individual, of which 92% were already annotated in a public database (dbSNP v129) (Table 2a). The proportion of previously annotated cSNPs was consistent by population, and higher for European (94%; $n = 6$) and Asian (93%; $n = 2$) than Yoruba (88%; $n = 4$) ancestry. These confirmation rates are $\sim 10\%$ higher than recent whole genome analyses^{12,19–22}. The most likely explanation is that coding sequences have historically been more heavily ascertained than noncoding sequences, although other factors such as dbSNP version, prior ascertainment of HapMap individuals and different false discovery rates may contribute as well. For the subset of cSNPs at coordinates with sufficient coverage for variant calling in all 12 individuals

($n = 47,079$), 32% of annotated variants and 86% of novel variants were singleton observations across 24 chromosomes (Fig. 1a).

We also estimated the total number of cSNPs in each individual relative to the reference genome (Table 2b). As the precise and comprehensive definition of the human exome remains incomplete, we extrapolated our data to an estimated exome size of exactly 30 Mb. The results were remarkably consistent by population. As expected, a higher number of non-synonymous cSNPs were estimated for the Yoruba individuals (average 10,254; $n = 4$) than non-Africans (average 8,489; $n = 8$). More heterozygous cSNPs were estimated for the four Yoruba (average 14,995) than the six European Americans (average 11,586) and the two Asians (average 10,631). The ratio of synonymous to non-synonymous cSNPs was 1.2 within any single individual, and 1.1 when calculated for a non-redundant list of variants identified across all individuals. The difference results from the slightly shifted allele frequency distribution of non-synonymous variants (Fig. 1b). Consistent with expectation²³, the trend is more pronounced for non-synonymous variants predicted to be damaging (by PolyPhen²⁴) (Fig. 1c).

Table 2 | Coding variation across 12 human exomes**a** Summary statistics for observed cSNPs

Individual	cSNP calls	Number in dbSNP	Percentage in dbSNP	Number heterozygous	Number homozygous
NA18507 (YRI)	19,720	17,577	89.1	12,896	6,824
NA18517 (YRI)	19,737	17,326	87.8	13,039	6,698
NA19129 (YRI)	19,761	17,298	87.5	12,845	6,916
NA19240 (YRI)	19,517	17,168	88.0	12,866	6,651
NA18555 (CHB)	16,047	14,894	92.8	9,181	6,866
NA18956 (JPT)	16,011	14,848	92.7	9,132	6,879
NA12156 (CEU)	16,119	15,250	94.6	10,179	5,940
NA12878 (CEU)	15,970	15,051	94.2	9,928	6,042
FSS10066 (Eur)	16,229	15,144	93.3	10,240	5,989
FSS10208 (Eur)	16,073	15,018	93.4	9,966	6,107
FSS22194 (Eur)	16,094	15,128	94.0	10,005	6,089
FSS24895 (Eur)	15,986	15,027	94.0	9,920	6,066

b Genome-wide cSNP estimates assuming a 30 Mb exome

Individual	Estimated total cSNPs	Estimated total heterozygous	Estimated total homozygous	Estimated total synonymous	Estimated total non-synonymous
NA18507 (YRI)	22,727	14,876	7,851	12,466	10,261
NA18517 (YRI)	22,841	15,135	7,706	12,550	10,291
NA19129 (YRI)	22,907	14,906	8,001	12,693	10,214
NA19240 (YRI)	22,814	15,063	7,751	12,565	10,249
NA18555 (CHB)	18,722	10,677	8,045	10,275	8,447
NA18956 (JPT)	18,523	10,585	7,938	10,072	8,451
NA12156 (CEU)	18,825	11,818	7,007	10,220	8,605
NA12878 (CEU)	18,544	11,455	7,089	10,110	8,434
FSS10066 (Eur)	18,836	11,795	7,041	10,240	8,596
FSS10208 (Eur)	18,591	11,444	7,147	10,075	8,516
FSS22194 (Eur)	18,667	11,539	7,128	10,144	8,523
FSS24895 (Eur)	18,508	11,466	7,042	10,169	8,339

For part **a**, cSNPs called in each individual, relative to the reference genome, are broken down by the fraction in dbSNP and by genotype. Part **b** shows extrapolation of observed numbers of cSNPs in each individual to an exactly 30 Mb exome. CEU, CEPH HapMap; CHB, Chinese HapMap; Eur, European-American ancestry (non-HapMap); JPT, Japanese HapMap; YRI, Yoruba HapMap.

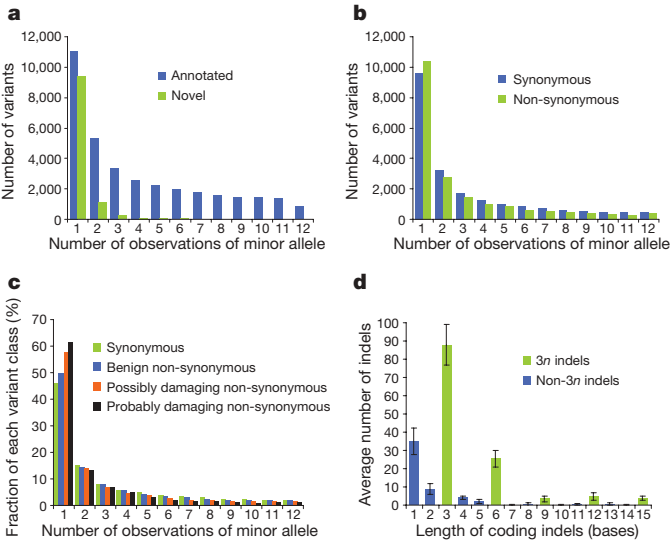


Figure 1 | Minor allele frequency and coding indel length distributions. **a**, The distribution of minor allele frequencies is shown for previously annotated versus novel cSNPs. **b**, The distribution of minor allele frequencies is shown for synonymous versus non-synonymous cSNPs. **c**, The distribution of minor allele frequencies (by proportion, rather than count) is shown for synonymous cSNPs ($n = 21,201$) versus non-synonymous cSNPs predicted to be benign ($n = 13,295$), possibly damaging ($n = 3,368$), or probably damaging ($n = 2,227$) by PolyPhen²⁴. **d**, The distribution of lengths of coding indel variants is shown (average numbers per exome). Error bars indicate s.d.

Nonsense mutations and splice-site disruptions are often assumed to be deleterious, but have a broad range of potential fitness effects^{25–27}. Our non-redundant cSNP catalogue included 225 nonsense mutations (112 novel) and 102 splice-site disruptions (49 novel). Excluding 86 nonsense alleles that are common in this data set (two or more observations) or in a recent study²⁵ ($>5\%$ allele frequency), our genome-wide estimate (projected to 30 Mb) for the average number of relatively rare mutations introducing premature nonsense codons in an individual genome was 10 for non-Africans ($n = 8$) and 20 for Yoruba ($n = 4$). However, these are probably overestimates, given that our catalogue of common nonsense mutations remains incomplete.

Short insertions and deletions (indels) in coding sequence are likely to be functionally important when they cause frameshifts, but are difficult to detect with short reads. We developed and applied an approach for identifying indels from our unpaired 76 bp reads. In total, 664 coding indels were called in one or more individuals. On average, 166 coding indels were called per individual, of which 63% were previously annotated in dbSNP (Supplementary Table 3). To assess our sensitivity, we compared our data for NA18507 to data published previously¹². The majority (73%) of their coding indels were also observed in our data (136 of 187). To assess specificity, we attempted PCR and Sanger sequencing of 28 novel coding indels chosen at random. Of 21 successful assays, 20 coding indels were verified and 1 was a false positive. We anticipate that future use of paired-end reads will improve detection of coding indels.

The shape of the distribution of coding indel lengths was consistent with other studies^{10,20} as well as across the 12 exomes (Fig. 1d), demonstrating a preference for multiples of 3 ($3n$). Of the 664 coding indels observed here, 65% were $3n$ in length. The allele frequency distribution for novel indels relative to annotated indels was markedly shifted towards rarer variants (Supplementary Fig. 4). However, the length histograms for novel versus annotated coding indels were similar (Supplementary Fig. 5), reinforcing the notion that our set of novel coding indels is not excessively contaminated with false positives (as these would not be expected to have the observed $3n$ bias). Excluding indels that were common in this data

set (two or more observations), the average number of relatively rare frameshifting indels identified per individual was 8 for non-Africans ($n = 8$) and 17 for Yoruba ($n = 4$).

The number of synonymous, missense, nonsense, splice site, frameshifting indel and non-frameshifting indel variants observed in each individual (as well as the size of the subsets that are novel and singleton observations) is presented in Supplementary Table 4. Also shown are the average numbers of variants of each class for non-Africans and Yoruba.

Phenotypes inherited in an apparently Mendelian pattern often lack sufficiently sized pedigrees to pinpoint the causal locus. We evaluated whether exome sequencing could be applied to identify directly the causative gene underlying a monogenic human disease (FSS), that is, with neither linkage data nor candidate gene analysis. Even in this simple scenario for ‘whole exome/genome genetics’, the key challenge that arises immediately is that the large number of apparently private mutations present by chance in any single human genome makes it difficult to identify which variant is causal, even when only considering non-synonymous variants. This hurdle was overcome recently in the context of hereditary pancreatic cancer by restricting focus only to nonsense mutations and also resequencing tumour DNA from the same individual, but this approach greatly limits sensitivity and is only relevant to a subset of mechanisms within one disease class²⁸.

To quantify this background of non-causal variants in our exome data, we first investigated how many genes had one or more non-synonymous cSNPs, splice site disruptions or coding indels in one or several FSS exomes (Fig. 2, row 1). Simply requiring that a gene contain variants in multiple affected individuals was clearly insufficient, as over 2,000 candidate genes remained even after intersecting four FSS exomes. We then applied filters to remove presumably common variants, as these are unlikely to be causative. Removing dbSNP-catalogued variants from consideration reduced the number of candidates considerably (Fig. 2, row 2). Remarkably, the eight HapMap exomes provided a filter nearly equivalent to dbSNP (Fig. 2, row 3). Combining the two catalogues had a synergistic effect (Fig. 2, row 4), such that the candidate list could be narrowed to a single gene (*MYH3*, identified previously by a candidate gene approach as causative for FSS⁵). Specifically, *MYH3* is the only gene where: (1) at least one (but not necessarily the same) non-synonymous cSNP, splice-site disruption or coding indel is observed in all four individuals with FSS; (2) the mutations are not in dbSNP, nor in the eight HapMap exomes. Taking the predicted deleteriousness of individual mutations into account served as an effective filter as well (Fig. 2, row 5), but was not required to identify *MYH3*. Ranges

Number of genes in which each affected has at least one...	Any 3 of 4				
	FSS24895	FSS24895 FSS10208	FSS24895 FSS10066	FSS24895 FSS10208 FSS10066	FSS24895 FSS10208 FSS10066 FSS22194
Non-synonymous cSNP, splice site variant or coding indel (NS/SS/I)	4,510	3,284	2,765	2,479	3,768
NS/SS/I not in dbSNP	513	128	71	53	119
NS/SS/I not in eight HapMap exomes	799	168	53	21	160
NS/SS/I neither in dbSNP nor eight HapMap exomes	360	38	8	1 (<i>MYH3</i>)	22
...And predicted to be damaging	160	10	2	1 (<i>MYH3</i>)	3

Figure 2 | Direct identification of the causal gene for a monogenic disorder by exome sequencing. Boxes list the number of genes with one or more non-synonymous cSNP, splice-site SNP, or coding indel (NS/SS/I) meeting specified filters. Columns show the effect of requiring that one or more NS/SS/I variants be observed in each of one to four affected individuals. Rows show the effect of excluding from consideration variants found in dbSNP, the eight HapMap exomes, or both. Column five models limited genetic heterogeneity or data incompleteness by relaxing criteria such that variants need only be observed in any three of four exomes for a gene to qualify.

of candidate list sizes when other permutations of individuals are used are shown in Supplementary Fig. 6.

MYH3 was well covered in our data. To assess our sensitivity more globally, we calculated the probability that a mutation would have been identified in all four FSS-affected individuals for each gene, based on our overall coverage of that gene in each individual (Supplementary Data 2). The average probability across all genes was 86%. This is probably still an overestimate of sensitivity, as functional noncoding or structural mutations would be missed. It also remains challenging to detect mutations in segmentally duplicated regions of the genome with short read sequencing.

Nevertheless, our analysis suggests that direct sequencing of exomes of small numbers of unrelated individuals (but more than one) with a shared monogenic disorder can serve as a genome-wide scan for the causative gene. The availability of the eight HapMap exomes was clearly helpful, suggesting that the power of this approach will improve as the 1000 Genomes Project²⁹ generates a catalogue of common variation that is more complete and evenly ascertained than dbSNP. Also, FSS is inherited in an autosomal dominant pattern so the presence of only one mutant allele is sufficient to cause disease. Applying this strategy to a recessive disease would probably be easier, because there are far fewer genes in each exome that are homozygous or compound heterozygous for rare non-synonymous variants. We also note that modelling of even a modest degree of genetic heterogeneity or data incompleteness is observed to have a significant impact on performance (Fig. 2, column offset to the right). Moving along the spectrum from rare monogenic disorders to complex common diseases, it is likely that the increasing extent of genetic heterogeneity will need to be matched by increasingly large sample sizes³⁰, and/or more sophisticated weighting of predicted mutational impact.

A clear limitation of exome sequencing is that it does not identify the structural and noncoding variants found by whole genome sequencing. At the same time, it allows a given amount of sequencing to be extended across at least 20 times as many samples compared to whole genome sequencing. In studies focused on identifying rare variants or somatic mutations with medical relevance, sample size and the interpretability of functional impact may be critical to achieving meaningful success. It is in the context of such studies that exome sequencing may be most valuable.

We demonstrate that targeted capture and massively parallel sequencing represents a cost-effective, reproducible and robust strategy for the sensitive and specific identification of variants causing protein-coding changes in individual human genomes. The 307 Mb determined here across 12 individuals is the largest data set reported so far of human coding sequence ascertained by second-generation sequencing methods. Finally, our successful demonstration that the causative gene for a Mendelian disorder can be identified directly by exome sequencing of several unrelated individuals provides increasing context to the possibility that exome or genome sequencing may represent a new approach for identifying gene–disease relationships.

METHODS SUMMARY

DNA samples, targeted capture and massively parallel sequencing. DNA samples were obtained from Coriell Repositories (HapMap) or by M.B. (FSS). Each shotgun library was hybridized to two Agilent 244K microarrays for target enrichment, followed by washing, elution and additional amplification. The first array targeted CCDS (2007), while the second was designed against targets poorly captured by the first array plus updates to CCDS in 2008. All sequencing was performed on the Illumina GA2 platform. Oligonucleotides used are listed in Supplementary Table 5.

Read mapping and variant analysis. Reads were mapped to the reference human genome (hg18, downloaded from <http://genome.ucsc.edu>), initially with ELAND (Illumina) for quality recalibration, and then again with Maq¹³. Sequence calls were also performed by Maq, and filtered to coordinates with $\geq 8\times$ coverage and a Phred-like¹⁵ consensus quality ≥ 30 . Sequence calls for HapMap individuals were compared against Illumina Human1M-Duo genotypes. NA18507 SNPs from whole genome data¹² were obtained from

Illumina. Annotations of cSNPs were based on NCBI and UCSC databases, supplemented with PolyPhen Grid Gateway²⁴ predictions for non-synonymous SNPs.

Identification of coding indels. Identification of coding indels involved: (1) gapped alignment of unmapped reads to the genome to generate a set of candidate indels using cross_match; (2) ungapped alignment of all reads to the reference and alternative alleles for all candidate indels using Maq; and (3) filtering by coverage and allelic ratio.

Data access. Sequencing reads for HapMap individuals are available from the NCBI Short Read Archive, accession SRP000910. Variants identified in HapMap individuals have been submitted to NCBI dbSNP under the handle 'SEATTLESEQ'. Variants identified in FSS individuals are available to approved investigators through NCBI dbGaP, accession number phs000204. Individual genotypes for variants identified in HapMap individuals, as well as the collapsed CCDS 2008 definition (before masking of coordinates listed in Supplementary Data 1), are available at http://krishna.gs.washington.edu/12_exomes.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 June; accepted 29 June 2009.

Published online 16 August 2009.

- Cohen, J. C. *et al.* Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305**, 869–872 (2004).
- Frazer, K. A., Murray, S. S., Schork, N. J. & Topol, E. J. Human genetic variation and its contribution to complex traits. *Nature Rev. Genet.* **10**, 241–251 (2009).
- Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nature Biotechnol.* **26**, 1135–1145 (2008).
- The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
- Toydemir, R. M. *et al.* Mutations in embryonic myosin heavy chain (*MYH3*) cause Freeman-Sheldon syndrome and Sheldon-Hall syndrome. *Nature Genet.* **38**, 561–565 (2006).
- Sjblom, T. *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268–274 (2006).
- Olson, M. Enrichment of super-sized resequencing targets from the human genome. *Nature Methods* **4**, 891–892 (2007).
- Hodges, E. *et al.* Genome-wide *in situ* exon capture for selective resequencing. *Nature Genet.* **39**, 1522–1527 (2007).
- National Center for Biotechnology Information. Consensus CDS protein set <<http://www.ncbi.nlm.nih.gov/projects/CCDS>> (2009).
- Ng, P. C. *et al.* Genetic variation in an individual human exome. *PLoS Genet.* **4**, e1000160 (2008).
- Kidd, J. M. *et al.* Mapping and sequencing of structural variation from eight human genomes. *Nature* **453**, 56–64 (2008).
- Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
- Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18**, 1851–1858 (2008).
- Campbell, P. J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature Genet.* **40**, 722–729 (2008).
- Ewing, B. & Green, P. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
- Turner, E. H., Lee, C., Ng, S. B. & Shendure, J. Massively parallel exon capture and library-free resequencing across 16 individuals. *Nature Methods* **6**, 315–316 (2009).
- Kidd, J. M. *et al.* Haplotype sorting using human fosmid clone end-sequence pairs. *Genome Res.* **18**, 2016–2023 (2008).
- Albert, T. J. *et al.* Direct selection of human genomic loci by microarray hybridization. *Nature Methods* **4**, 903–905 (2007).
- Wheeler, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872–876 (2008).
- Wang, J. *et al.* The diploid genome sequence of an Asian individual. *Nature* **456**, 60–65 (2008).
- Levy, S. *et al.* The diploid genome sequence of an individual human. *PLoS Biol.* **5**, e254 (2007).
- Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66–72 (2008).
- Boyko, A. R. *et al.* Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.* **4**, e1000083 (2008).
- Sunyaev, S. *et al.* Prediction of deleterious human alleles. *Hum. Mol. Genet.* **10**, 591–597 (2001).
- Yngvadottir, B. *et al.* A genome-wide survey of the prevalence and evolutionary forces acting on human nonsense SNPs. *Am. J. Hum. Genet.* **84**, 224–234 (2009).
- Olson, M. V. When less is more: gene loss as an engine of evolutionary change. *Am. J. Hum. Genet.* **64**, 18–23 (1999).
- Cohen, J. *et al.* Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nature Genet.* **37**, 161–165 (2005).

28. Jones, S. *et al.* Exomic sequencing identifies *PALB2* as a pancreatic cancer susceptibility gene. *Science* **324**, 217 (2009).
29. Siva, N. 1000 Genomes project. *Nature Biotechnol.* **26**, 256 (2008).
30. Kryukov, G. V., Shpunt, A., Stamatoyannopoulos, J. A. & Sunyaev, S. R. Power of deep, all-exon resequencing for discovery of human trait genes. *Proc. Natl Acad. Sci. USA* **106**, 3871–3876 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements For discussions or assistance with genotyping data, we thank P. Green, J. Akey, R. Patwardhan, G. Cooper, J. Kidd, D. Gordon, J. Smith, I. Stanaway and M. Rieder. For assistance with project management, computation, data management and submission, we thank E. Torskey, S. Thompson, T. Amburg, B. McNally, S. Hearsey, M. Shumway and L. Hillier. For Human1M-Duo genotype data on HapMap samples, we thank Illumina. Our work was supported in part by grants from the National Institutes of Health/National Heart Lung and Blood Institute, the National Institutes of Health/National Human Genome Research

Institute, National Institutes of Health/National Institute of Child Health and Human Development, and the Washington Research Foundation. S.B.N. is supported by the Agency for Science, Technology and Research, Singapore. E.H.T. and A.W.B. are supported by a training fellowship from the National Institutes of Health/National Human Genome Research Institute. E.E.E. is an investigator of the Howard Hughes Medical Institute.

Author Contributions The project was conceived and experiments planned by S.B.N., E.H.T., A.B., E.E.E., M.B., D.A.N. and J.S. Experiments were performed by S.B.N., E.H.T., C.L. and M.W. Algorithm development and data analysis were performed by S.B.N., P.D.R., S.D.F., A.W.B., T.S., M.B., D.A.N. and J.S. The manuscript was written by S.B.N. and J.S. All aspects of the study were supervised by J.S.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.S. (shendure@u.washington.edu) or S.B.N. (sarahng@u.washington.edu).

METHODS

Genomic DNA samples. Targeted capture was performed on genomic DNA from eight HapMap individuals (four Yoruba (NA18507, NA18517, NA19129 and NA19240), two East Asians (NA18555 and NA18956) and two European-Americans (NA12156 and NA12878)) and four European-American individuals affected by Freeman–Sheldon syndrome (FSS10066, FSS10208, FSS22194 and FSS24895). Genomic DNA for HapMap individuals was obtained from Coriell Cell Repositories. Genomic DNA for FSS individuals was obtained by M.B.

Oligonucleotides and adaptors. All oligonucleotides were synthesized by Integrated DNA Technologies and resuspended in nuclease-free water to a stock concentration of 100 μ M. Sequences are shown in Supplementary Table 5. Double-stranded library adaptors SLXA_1 and SLXA_2 were prepared to a final concentration of 50 μ M by incubating equimolar amounts of SLXA_1_HI and SLXA_1_LO together and SLXA_2_HI and SLXA_2_LO together at 95 °C for 3 min and then leaving the adaptors to cool to room temperature in the heat block.

Shotgun library construction. Shotgun libraries were generated from 10 μ g of genomic DNA (gDNA) using protocols modified from the standard Illumina protocol¹². Each library provided sufficient material for hybridization to two microarrays. For each sample, gDNA in 300 μ l 1 \times Tris-EDTA was first sonicated for 30 min using a Bioruptor (Diagenode) set at high, then end-repaired for 45 min in a 100 μ l reaction volume using 1 \times End-It Buffer, 10 μ l dNTP mix and 10 μ l ATP as supplied in the End-It DNA End-Repair Kit (Epicentre). The fragments were then A-tailed for 20 min at 70 °C in a 100 μ l reaction volume with 1 \times PCR buffer (Applied Biosystems), 1.5 mM MgCl₂, 1 mM dATP and 5 U AmpliTaq DNA polymerase (Applied Biosystems). Next, library adaptors SLXA_1 and SLXA_2 were ligated to the A-tailed sample in a 90 μ l reaction volume with 1 \times Quick Ligation Buffer (New England Biolabs) with 5 μ l Quick T4 DNA Ligase (New England Biolabs) and each adaptor in 10 \times molar excess of sample. Samples were purified on QIAquick columns (Qiagen) after each of these four steps and DNA concentration determined on a Nanodrop-1000 (Thermo Scientific) when necessary.

Each sample was subsequently size-selected for fragments of size 150–250 bp using gel electrophoresis on a 6% TBE-polyacrylamide gel (Invitrogen). A gel slice containing the fragments of interest was then excised and transferred to a siliconized 0.5 ml microcentrifuge tube (Ambion) with a 20 G needle-punched hole in the bottom. This tube was placed in a 1.5 ml siliconized microcentrifuge tube (Ambion), and centrifuged in a tabletop microcentrifuge at 16,110g for 5 min to create a gel slurry that was then resuspended in 200 μ l 1 \times Tris-EDTA and incubated at 65 °C for 2 h, with periodic vortexing. This allowed for passive elution of DNA, and the aqueous phase was then separated from gel fragments by centrifugation through 0.2 μ m NanoSep columns (Pall Life Sciences) and the DNA recovered using a standard ethanol precipitation.

Recovered DNA was resuspended in elution buffer (EB; 10 mM Tris-Cl, pH 8.5, Qiagen) and the entire volume used in a 1 ml bulk PCR reaction volume with 1 \times iProof High-Fidelity Master Mix (Bio-Rad) and 0.5 μ M each of primers SLXA_FOR_AMP and SLXA_REV_AMP with the conditions: 98 °C for 30 s, 20 cycles at 98 °C for 30 s, 65 °C for 10 s and 72 °C for 30 s, and finally 72 °C for 5 min. PCR products were purified across four QIAquick columns (Qiagen) and all the eluants pooled.

Design of exome capture arrays. We targeted all well-annotated protein-coding regions as defined by the CCDS (version 20080902). Coordinates were extracted from entries with ‘public’ status, and regions with overlapping coordinates were merged. This resulted in a target with 164,007 discontinuous regions summing to 27,931,548 bp. By comparison, coding sequence defined by all of RefSeq (NCBI 36.3) comprises 31.9 Mb (14% larger). Hybridization probes against the target were designed primarily such that they were evenly spaced across each region. Probes were also constrained (1) to be relatively unique, such that the average occurrence of each 15-mer in the probe sequence is less than 100⁸, (2) to be between 20 and 60 bases in length, with preference for longer probes, and (3) to have a calculated melting temperature (T_m) \leq 69 °C, with preference for higher T_m values. T_m was calculated by $64.9 + 41 \times (\text{number of G + Cs} - 16.4)/\text{length of probe}$.

Two arrays (Agilent, 244K format) were designed and used per individual. The first array was common to all individuals, and contained 241,071 probes designed mainly against the subset of the target that was also found in a previous version of the CCDS (CCDS20070227). For most exomes, the second array was custom-designed specifically against target regions that had not been adequately represented after capture on the first array and subsequent sequencing. For two individuals (FSS10066, FSS10208), the matching was to a different individual’s first-array data. However, this did not seem to have a significant effect on performance, probably because features capturing poorly on the first array largely did so consistently. Additionally, all of the second arrays also targeted sequences

found in CCDS20080902 that were not in CCDS20070227 and hence not targeted by the first array. A subset of arrays used lacked control grids.

Targeted capture by hybridization to DNA microarrays. Hybridizations to Agilent 244K arrays were performed following manufacturer’s instructions with modifications. For each enrichment, a 520 μ l hybridization solution containing 20 μ g of the bulk-amplified genomic DNA library, 1 \times aCGH hybridization buffer (Agilent), 1 \times blocking agent (Agilent), 50 μ g human CotI DNA (Invitrogen) and 0.92 nmol each of the blocking oligonucleotides SLXA_FOR_AMP, SLXA_REV_AMP, SLXA_FOR_AMP_rev and SLXA_REV_AMP_rev was incubated at 95 °C for 3 min and then at 37 °C for at least 30 min. The hybridization solution was then loaded and the hybridization chamber assembled following the manufacturer’s instructions. Incubation was done at 65 °C for at least 66 h with rotation at 20 r.p.m. in a hybridization oven (Agilent).

After hybridization, the slide-gasket sandwich was removed from the chamber and placed in a 50 ml conical tube filled with aCGH Wash Buffer 1 (Agilent). The slide was separated from the gasket while in the buffer and then washed, first with fresh aCGH Wash Buffer 1 at room temperature for 10 min on an orbital shaker (VWR) set on low speed, and then in pre-warmed aCGH Wash Buffer 2 (Agilent) at 37 °C for 5 min. Both washes were also done in 50 ml conical tubes.

A Secure-Seal (SA2260, Grace Bio Labs) was then affixed firmly over the active area of the washed slide and heated briefly according to the manufacturer’s instructions. One port was sealed with a seal tab and the seal chamber completely filled with approximately 1 ml of hot EB (95 °C). The other port was sealed and the slide incubated at 95 °C on a heat block. After 5 min, one port was unsealed and the solution recovered. DNA was purified from the solution using a standard ethanol precipitation.

Precipitated DNA was resuspended in EB and the entire volume used in a 50 μ l PCR volume comprising of 1 \times iTaq SYBR Green Supermix with ROX (Bio-Rad) and 0.2 μ M each of primers SLXA_FOR_AMP and SLXA_REV_AMP. Thermal cycling was done in a MiniOpticon Real-time PCR system (Bio-Rad) with the following programme: 95 °C for 5 min, then 30 cycles of 95 °C for 30 s, 55 °C for 2 min and 72 °C for 2 min. Each sample was monitored and extracted from the PCR machine when fluorescence began to plateau. Samples were then purified on a QIAquick column (Qiagen) and sequenced.

Sequencing. All sequencing of post-enrichment shotgun libraries was carried out on an Illumina Genome Analyzer II as single-end 76 bp reads, following the manufacturer’s protocols and using the standard sequencing primer. Image analysis and base calling was performed by the Genome Analyser Pipeline version 1.0 or 1.3 with default parameters, but with no pre-filtering of reads by quality. Quality values were recalibrated by alignment to the reference human genome with the Eland module.

Read mapping. The reference human genome used in these analyses was UCSC assembly hg18 (NCBI build 36.1), including unordered sequence (chrN_random.fa) but not including alternate haplotypes. For each lane, reads with calibrated qualities were extracted from the Eland export output. Base qualities were rescaled and reads mapped to the human reference genome using Maq (version 0.7.1)¹³. Unmapped reads were dumped using the $-u$ option and subsequently used for indel mapping. Mapped reads that overlapped target regions (‘target reads’) were used for all other analyses.

Target masking. All possible 76-bp reads that overlapped the aggregate target were simulated, mapped using Maq and consensus called using Maq assemble with parameters $-q 1 -r 0.2 -t 0.9$. Target coordinates that had read depth < 76 (that is, half of the expected depth), reflecting a poor ability to have reads confidently mapped to them (Supplementary Data 1), were removed from consideration for downstream analyses, leaving a 26,553,795 bp target.

Variant calling. All reads with a map score > 0 from each individual were merged and filtered for duplicates such that only the read with the highest aggregate base quality at any given start position and orientation was retained. Sequence calls were obtained using Maq assemble with parameters $-r 0.2 -t 0.9$, and only coordinates with at least 8 \times coverage and an estimated Phred-like consensus quality value of at least 30 were used for downstream variant analyses.

Comparison of sequence calls to array genotypes, dbSNP and whole genome sequencing. For the eight HapMap individuals, sequence calls were compared to array-based genotyping data (Illumina Human1M-Duo) provided by Illumina. We excluded from consideration genotyping assays where all eight individuals were called by the arrays as homozygous non-reference as well as the MHC locus at chromosome 6:32500001–33300000, as both sets are likely to be error-enriched in the genotyping data. We downloaded dbSNP(v129) from ftp://ftp.ncbi.nih.gov/snp/organisms/human_9606/chr_rpts on 13 May 2008. Approximately 14.2 million non-redundant coordinates were defined by this file set. For comparison of NA18507 cSNPs to whole genome data, variant lists were obtained from Illumina¹².

Identification of coding indels. Reads for which Maq was unsuccessful in identifying an ungapped alignment were converted to fasta format and mapped to the

human reference genome with `cross_match` (v1.080812, <http://www.phrap.org>), using parameters `-gap_ext -1 -bandwidth 10 -minmatch 20 -maxmatch 24`. Output options `-tags -discrep_lists -alignments -score_hist` were also set. Alignments with an indel were then filtered for those that: (1) had a score at least 40 more than the next best alignment; (2) mapped at least 75 bases of the read; (3) had no substitutions in addition to the indel; and (4) overlapped a target region. Reads from filtered alignments that mapped to the negative strand were then reverse-complemented and, together with the rest of the filtered reads, re-mapped with `cross_match` using the same parameters. This was to reduce ambiguity in called indel positions due to different read orientations. After the second mapping, alignments were re-filtered using the same criteria (1) to (4). For each sample, a putative indel event was called if at least two filtered reads covered the same event. A fasta file containing the sequences of all called events ± 75 bp, as well as the reference sequence at the same positions, was then generated for each individual. All the reads from each individual were then mapped to its 'indel reference' with `Maq` using default parameters. Reads that mapped multiple times (map score 0) or had redundant start sites were removed, after which the number of reads mapping to either the reference or the non-reference allele was counted for each individual and indel. An indel was called if there were at least eight non-reference allele reads making up at least 30% of all reads at that genomic position. Indels were called as heterozygous if non-reference alleles were 30–70% of reads at that position, and homozygous non-reference if $>70\%$.

Variant annotation. For cSNP annotation, we constructed a local server that integrates data from NCBI (including dbSNP and Consensus CDS files) and from UCSC Genome Bioinformatics. We also generated PolyPhen predictions²⁴ for all cSNPs identified here, using the PolyPhen Grid Gateway and Perl scripts

supplied by I. Adzhubey. The server reads files with SNP locations and alleles, and produces annotation files available for download. Annotation includes dbSNP rs IDs, overlapping-gene accession numbers, SNP function (for example, whether coding missense), conservation scores, HapMap minor-allele frequencies and various protein annotations (sequence, position, amino acid changes with physicochemical properties and PolyPhen classification). Indels were considered annotated by dbSNP if an entry was found with the same allele (or reverse complemented) within 1 bp of the variant position. This was to allow for ambiguities in calling the indel position.

Calculation of genome-wide estimates. Extrapolated estimates for the genome-wide number of cSNPs of various classes (Table 2b) were calculated based on the number of cSNP calls in that individual, the estimated sensitivity for making a variant call in that individual at any given position within the aggregate target (based on the fraction of array-based genotypes of that class that were successfully called; calculated separately for heterozygous and homozygous non-reference variants), and extrapolation to an estimated exome size of exactly 30 Mb (that is, multiplying by $30/26.6 = 1.13$). A similar approach was taken to estimate the genome-wide number of uncommon cSNPs introducing non-sense codons, starting with the number observed in each individual and extrapolating based on estimated sensitivity for heterozygote detection and an estimated exome size of exactly 30 Mb.

Freeman–Sheldon syndrome mutations. For FSS10066, FSS22194 and FSS24895, the identified mutation was a C→T at chromosome 17:10485359, and the corresponding amino acid change was R672H. For FSS10208, the mutation was C→T at chromosome 17:10485360, and the corresponding amino acid change was R672C.

Modification of CO₂ avoidance behaviour in *Drosophila* by inhibitory odorants

Stephanie Lynn Turner¹ & Anandasankar Ray²

The fruitfly *Drosophila melanogaster* exhibits a robust and innate olfactory-based avoidance behaviour to CO₂, a component of odour emitted from stressed flies¹. Specialized neurons in the antenna and a dedicated neuronal circuit in the higher olfactory system mediate CO₂ detection and avoidance^{1,2}. However, fruitflies need to overcome this avoidance response in some environments that contain CO₂ such as ripening fruits and fermenting yeast, which are essential food sources. Very little is known about the molecular and neuronal basis of this unique, context-dependent modification of innate olfactory avoidance behaviour. Here we identify a new class of odorants present in food that directly inhibit CO₂-sensitive neurons in the antenna. Using an *in vivo* expression system we establish that the odorants act on the Gr21a/Gr63a CO₂ receptor³. The presence of these odorants significantly and specifically reduces CO₂-mediated avoidance behaviour, as well as avoidance mediated by 'Drosophila stress odour'. We propose a model in which behavioural avoidance to CO₂ is directly influenced by inhibitory interactions of the novel odours with CO₂ receptors. Furthermore, we observe differences in the temporal dynamics of inhibition: the effect of one of these odorants lasts several minutes beyond the initial exposure. Notably, animals that have been briefly pre-exposed to this odorant do not respond to the CO₂ avoidance cue even after the odorant is no longer present. We also show that related odorants are effective inhibitors of the CO₂ response in *Culex* mosquitoes that transmit West Nile fever and filariasis. Our findings have broader implications in highlighting the important role of inhibitory odorants in olfactory coding, and in their potential to disrupt CO₂-mediated host-seeking behaviour in disease-carrying insects like mosquitoes.

CO₂ is an important sensory cue for many animals, including insects, in a variety of behavioural contexts^{4,5}. In *Drosophila*, CO₂ is exclusively detected by a unique heteromeric receptor encoded by Gr21a and Gr63a (refs 3, 6–8) that is expressed in a single population of antennal olfactory receptor neurons (ORNs), called ab1C, which innervate the ab1 class of large basiconic sensilla^{1,7}. These neurons send stereotypical axonal projections to the V glomerulus, and activation of this dedicated uni-glomerular circuit leads to an innate avoidance of CO₂ (refs 1, 2, 9).

In fact, CO₂ is a major component of *Drosophila* stress odour (dSO), which is emitted by flies subjected to vigorous shaking or electric shock, and which elicits an immediate escape response in naive flies¹. However, CO₂ is also present in significant quantities in several important food sources that elicit behavioural attraction of *Drosophila*. Fruits and plants emit CO₂ as a by-product of respiration, as do fruits undergoing fermentation by microorganisms and yeasts^{4,9–11}. Flies are attracted to headspace odours containing CO₂ collected from over-ripe fruits, fermenting yeast and beer when presented with a choice between two tubes in a T-maze assay, one containing air and the other containing headspace odours

(Supplementary Fig. 1). However, flies avoid headspace odours collected from green fruits, which also emit CO₂ (ref. 9). A subset of specialized gustatory neurons mediate a small degree of attraction to carbonated water upon contact¹²; however, they do not respond to CO₂ in the gas phase and are not likely to contribute to long-range or short-range behavioural attraction towards a food source. Therefore, olfactory avoidance to CO₂ may be modified by context for some CO₂-rich sources such as over-ripe fruit, yeast and beer.

Little is known about the molecular and neuronal mechanisms that lead to such a dramatic modification of innate avoidance behaviour¹³. Two alternative models, although not mutually exclusive, may be evoked to explain this phenomenon. In the first model (Fig. 1a, top), avoidance to CO₂ is overcome simply by detection of attractive odorants emitted by the same food sources. In the second model (Fig. 1a, bottom), some components of food volatiles may also directly inhibit the CO₂-responsive circuit, and thereby suppress avoidance behaviour to CO₂.

To test whether odorants present in fruits and other natural environments of *Drosophila* can directly inhibit CO₂-sensitive ab1C neurons, we performed a simple electrophysiology screen. We tested several individual odorants for their ability to inhibit the baseline activity of the ab1C neuron (to about 0.03% CO₂ present in room air) using single-sensillum electrophysiology. We performed these experiments using *Or83b*² mutant flies in which the ab1C neuron remains the sole functional neuron in the ab1 sensillum^{14,15}. In a screen with 46 odorants, we identified two, 1-hexanol and 2,3-butanedione, that strongly inhibit the baseline activity of ab1C neurons (Fig. 1b). Both of these compounds are present in *Drosophila* food sources including various types of fruit (Supplementary Tables 1 and 2). More interestingly, the abundance of both these compounds is greatly increased during the ripening process of fruits: for example, in banana 1-hexanol increases by 777% and 2,3-butanedione by 14,900%¹⁶ (Supplementary Table 2). 1-Hexanol is formed during ripening by lipid oxidation of unsaturated fatty acids¹⁷, whereas 2,3-butanedione is a natural by-product of fermentation of carbohydrates through pyruvate by yeasts and bacteria and is thus also present in fermenting fruit, wine¹⁸ and beer^{19,20}.

We found that both 1-hexanol and 2,3-butanedione inhibit CO₂ response in a dose-dependent manner, irrespective of whether their application is initiated before, or after, the presentation of the CO₂ stimulus (Fig. 1c, d) at relatively low, physiologically relevant concentrations (see Methods and Supplementary Fig. 2).

A fly approaching an odour source from a distance likely contacts plumes of CO₂, which will vary widely in concentration over baseline atmospheric levels²¹. When we tested several concentrations of CO₂ we found that the presence of 2,3-butanedione (10^{−1} dilution) completely inhibits responses up to 3.2% CO₂ (Fig. 1e); 1-hexanol (10^{−1} dilution) also causes a significant reduction of CO₂ response across most tested concentrations, but complete inhibition occurs only at 0.1% CO₂ (Fig. 1e).

¹Cellular, Molecular, and Developmental Biology Program, ²Department of Entomology, University of California, Riverside, California 92521, USA.

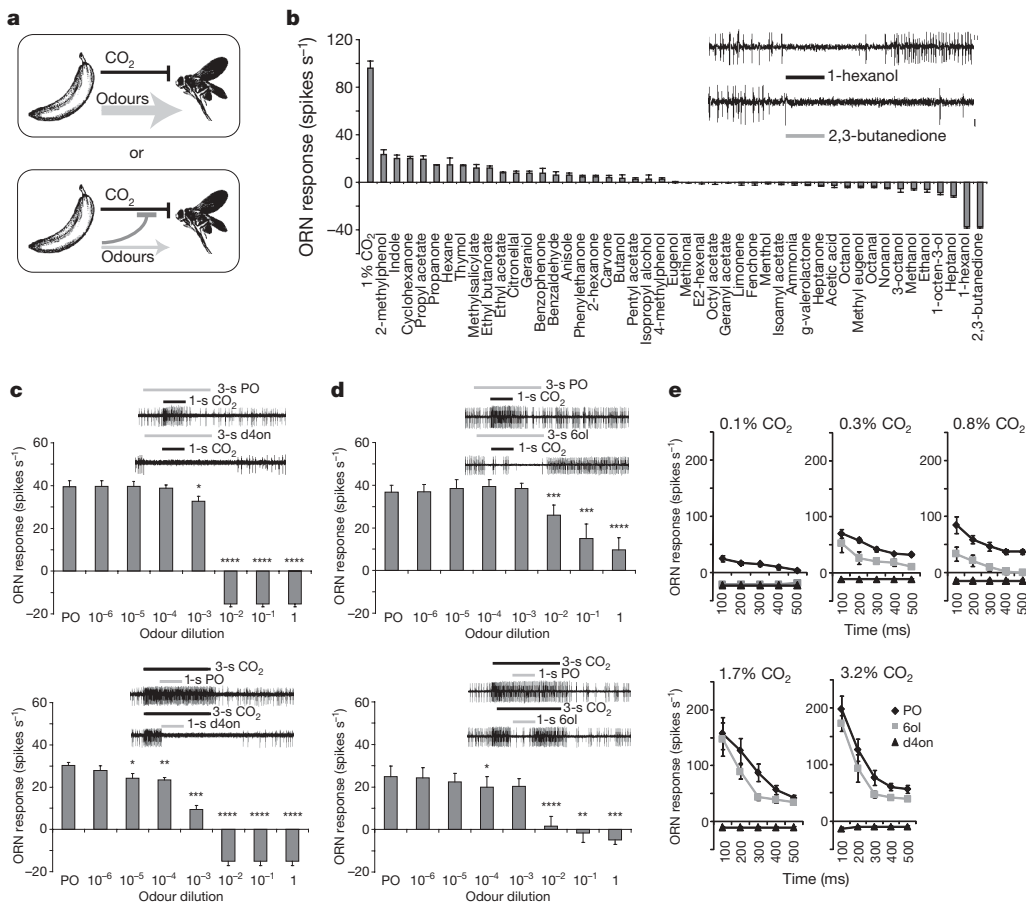


Figure 1 | Inhibitory odours dramatically reduce response to CO₂. **a**, Proposed models for suppression of avoidance behaviour to CO₂ in the context of fruits. **b**, Mean odorant responses and representative traces of activity of ab1C neurons using single-sensillum electrophysiology in *Or83b*² flies. Bars indicate a 0.5-s stimulus period. Odorants were tested at 10⁻² dilution in paraffin oil. Bars represent values after subtraction of mean response to paraffin oil ($n = 3$, error bars, s.e.m.). **c**, **d** Representative traces and mean responses from single-sensillum electrophysiology of ab1 sensilla in *Or83b*² flies; spikes and bars represent activity of the ab1C neuron. Top, 3-s stimulus of odorant overlaid with a 1-s application of 0.3% CO₂; bottom, 3-s stimulus of 0.3% CO₂ overlaid with a 1-s application of odorant (PO, paraffin oil; d4on, 2,3-butanedione; 6ol, 1-hexanol) ($n = 6$, error bars, s.e.m.). Spikes per second were counted during the 1-s stimulus period, and spontaneous activity subtracted. (For data in **c** and **d**, t -test, **** $P < 0.001$, *** $P < 0.005$, ** $P < 0.01$, * $P < 0.05$.) **e**, ab1C responses to indicated concentrations of CO₂ in the presence of solvent (PO), 1-hexanol (6ol) or 2,3-butanedione (d4on). Odorants were tested at 10⁻¹ dilution. Firing rates were counted in consecutive 0.1-s bins ($n = 5$, error bars, s.e.m.).

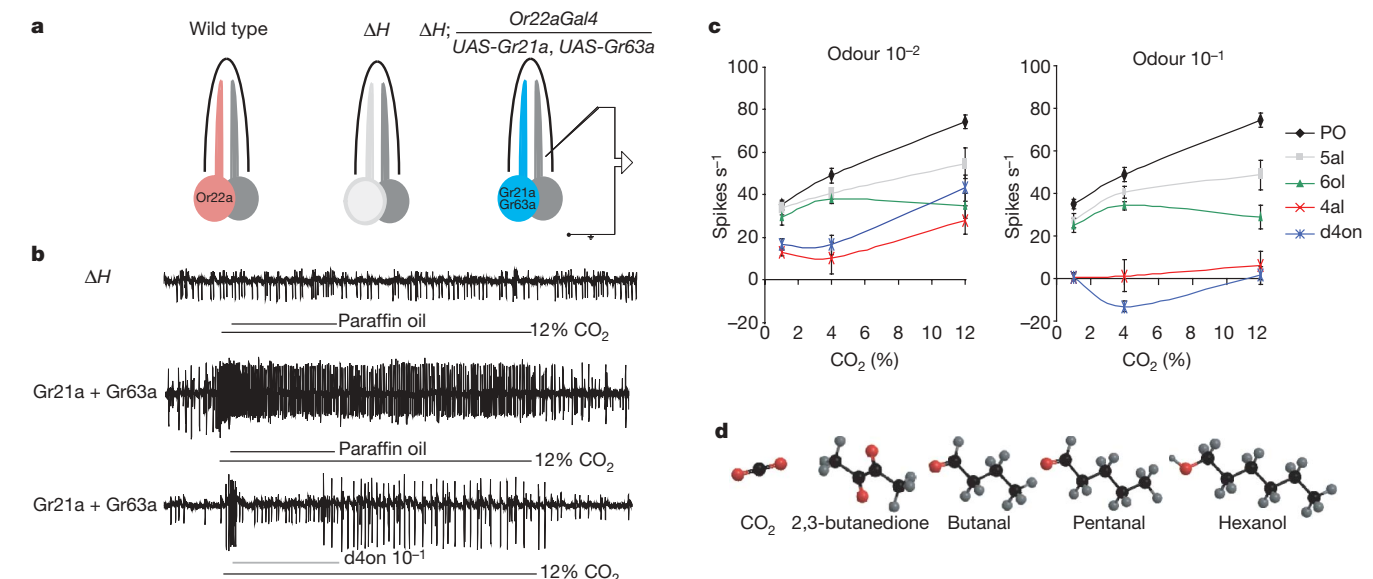


Figure 2 | Inhibitory odorants directly affect CO₂ response of Gr21a/Gr63a in a heterologous system. **a**, Schematic illustrating 'empty neuron' system used for heterologous expression of Gr21a and Gr63a in ab3A neurons. **b**, Representative traces of recordings from ab3 sensilla. Large spikes represent the response of the Δ ab3A cell expressing Gr21a and Gr63a. Bars indicate stimulus periods of 12% CO₂, overlaid with paraffin oil (PO) or

2,3-butanedione (d4on) at 10⁻¹ dilution. **c**, Concentration-dependent responses of ab3A neuron to CO₂, and binary mixtures of CO₂ with odorants at indicated concentrations: 1-butanol (4al), 1-pentanol (5al), 1-hexanol (6ol). Stimuli were applied as in **b** ($n = 5$, error bars, s.e.m.). **d**, Structures of CO₂ and the four strongest antagonists.

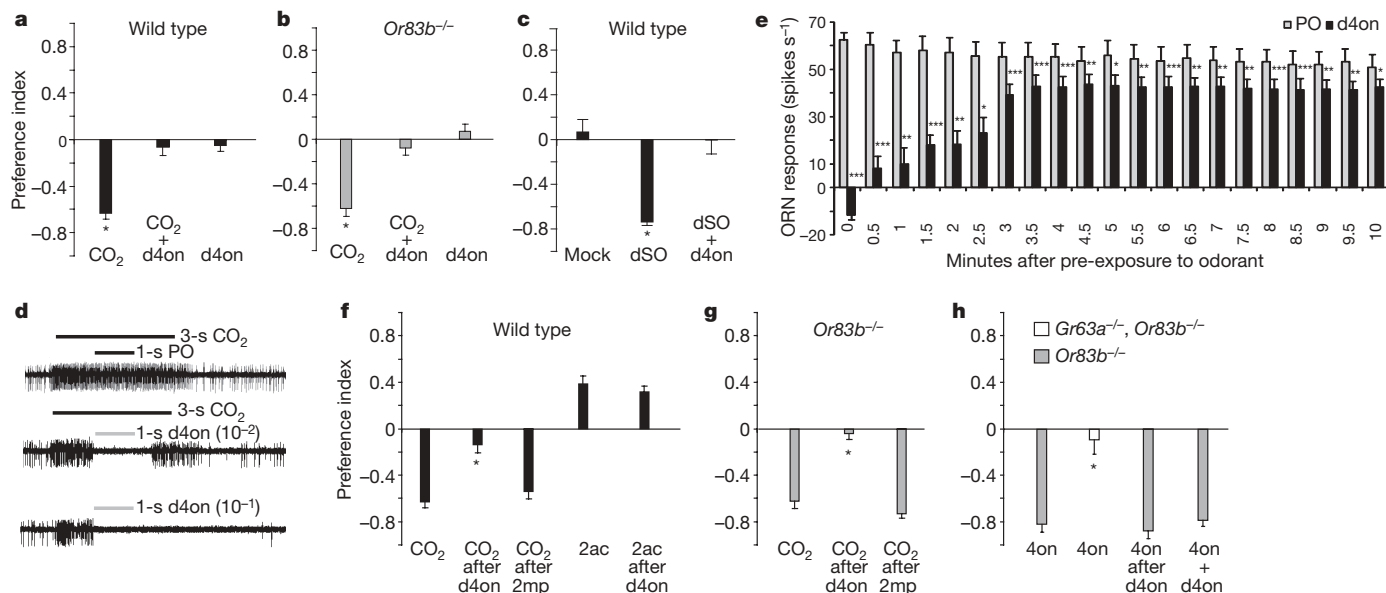


Figure 3 | Avoidance behaviour to CO₂ and *Drosophila* stress odour is abolished by inhibitory odorants. T-maze behaviour assay: **a**, mean preference index of wild-type flies, given a choice between room air in a 15-ml tube and either 0.1 ml of pure CO₂ (CO₂), a binary mixture of 0.1 ml pure CO₂ and 2,3-butanedione at 10⁻² dilution (CO₂ + d4on) or 2,3-butanedione at 10⁻² dilution (d4on) also in 15-ml tubes (see Methods). **b**, Mean preference index of *Or83b*^{-/-} mutant flies given choices as in **a**. **c**, Mean preference index of wild-type flies given choices between room air and odour collected from 70 untreated flies (mock), or dSO collected from 70 vortexed 'emitter' flies, or a mixture of dSO with d4on at 10⁻² dilution (dSO + d4on) (n = 6–9 trials (approximately 40 flies each); error bars, s.e.m. (t-test, *P < 0.0001)). **d**, Representative traces of ab1C neuronal activity. Bars indicate stimulus periods for 0.3% CO₂ overlaid with paraffin oil (PO) or 2,3-butanedione (d4on) at the indicated concentrations. **e**, Recovery of ab1C responsiveness to a 0.5-s, 0.3% CO₂ stimulus applied every 30 s after initial treatment with a 3-s stimulus of either d4on (10⁻¹ dilution) or paraffin oil

(PO) (n = 5; error bars, s.e.m. (t-test, ***P < 0.005, ** < 0.01, *P < 0.05)). **f**, T-maze behaviour assay: mean preference index of wild-type flies, given a choice between room air and CO₂ (CO₂) or ethyl acetate (2ac) (10⁻⁴ dilution). Experiments were performed as in **a**, or after a 1-min pre-exposure to either 2,3-butanedione (10⁻² dilution) (after d4on), or 2-methyl phenol (10⁻² dilution) (after 2mp) as indicated and a subsequent 2-min hold in clean air (n = 6 trials (approximately 40 flies each); error bars, s.e.m. (t-test, *P < 0.0001)). **g**, Mean preference index of *Or83b*^{-/-} mutant flies given choices as indicated, as described for **f** (n = 6 trials (approximately 40 flies each); error bars, s.e.m. (t-test, *P < 0.0001)). **h**, Mean preference index of *Or83b*^{-/-} or *Gr63a*^{-/-}, *Or83b*^{-/-} flies as indicated, assayed as in **f**, given a choice between room air and butanone (4on) (10⁻¹ dilution); or *Or83b*^{-/-} flies given the same choice, after a 1-min pre-exposure to d4on (10⁻² dilution) as in **g** (4on after d4on), or in the presence of 2,3-butanedione (10⁻² dilution) (4on + d4on) (n = 6 trials (20 flies each); error bars, s.e.m. (t-test, *P < 0.0001)).

To understand odorant structural features that might have a role in inhibition, we tested a rationally designed panel of odorants that varied in the number of carbon atoms and in the nature of the functional group. On the basis of this analysis, we identified additional odorants that also inhibit CO₂ response (Supplementary Figs 3 and 4a). The inhibitory effects of each of the compounds we have identified so far are specific to the CO₂-sensitive neuron; previous studies have shown that all of them can excite other classes of *Drosophila* ORNs^{22,23}, which suggests that they are not general inhibitors of ORN function. Surprisingly, these compounds are structurally quite different from CO₂ (Fig. 2d), thus raising the possibility that they

may act through allosteric binding sites within the Gr21a/Gr63a receptor, or on other components of the CO₂ detection pathway such as factors present in the sensillar lymph or in ab1C neurons.

To investigate whether the inhibitors act directly on the CO₂ receptor, we expressed *Gr21a* and *Gr63a* in an *in vivo* decoder system called the 'empty neuron'^{3,8} (Fig. 2a). We found that expression of *Gr21a* and *Gr63a* in the empty ab3A neuron is sufficient to impart a robust and reproducible dose-dependent CO₂ response, comparable to the levels reported previously³ (Fig. 2b, c). Upon application of each of the four inhibitory odorants along with CO₂, we observed dose-dependent inhibition of CO₂ response of the ab3A neuron

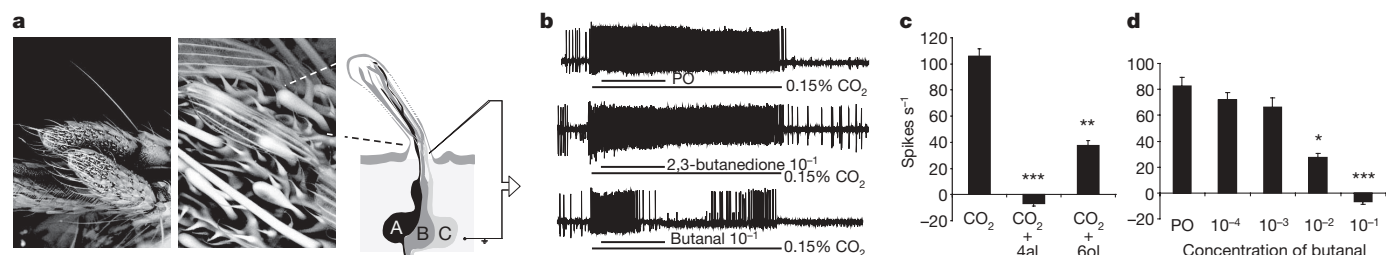


Figure 4 | Inhibitory odours dramatically reduce CO₂ response in *Culex quinquefasciatus* mosquitoes. **a**, Scanning electron micrographs of female *C. quinquefasciatus* maxillary palps (left) and peg sensilla (centre). Schematic of a peg sensillum containing three neurons (right). **b**, Representative traces from single-sensillum electrophysiology recordings of peg sensilla in female *C. quinquefasciatus*. The largest spike amplitude represents activity of the CO₂-sensitive A neuron. A 1-s stimulus of either

paraffin oil solvent (PO), 2,3-butanedione (d4on) or 1-butanol (4al) at 10⁻¹ dilution is overlaid over a 3-s stimulus of 0.15% CO₂. **c**, Mean responses of the A neuron to a 1-s stimulus of 1-butanol at indicated concentrations applied over a 3-s stimulus of 0.15% CO₂ (n = 5; error bars, s.e.m.). (For data in **c** and **d**, t-test, ***P < 0.000001, **P < 0.00001, *P < 0.00005).

(Fig. 2b, c) in a Gr21a/Gr63a-dependent manner (Supplementary Fig. 4b). The simplest interpretation of these results is that the odorants we have identified inhibit CO₂ response by direct interaction with the CO₂ receptor, Gr21a/Gr63a. However, the inhibitory effect appears shorter in duration than observed in the endogenous ab1C neurons, suggesting that additional neuron- or sensillum-specific factors may also influence the temporal aspects of the inhibition.

We next asked whether the inhibitory odorants identified using electrophysiology could disrupt avoidance behaviour of *Drosophila* to CO₂. Using a T-maze choice assay as described earlier¹, we found that wild-type *Drosophila* show a robust avoidance behaviour to 0.67% CO₂ (Fig. 3a). Inclusion of 2,3-butanedione with CO₂ completely abolishes avoidance to CO₂ (Fig. 3a). Importantly, 2,3-butanedione by itself does not elicit any significant attraction or avoidance behaviour (Fig. 3a). In wild-type *Drosophila*, however, several other ORN classes are activated by 2,3-butanedione^{22,23}, raising the possibility that behavioural avoidance to CO₂ may be overcome by activation of these other classes of ORN, and not solely by inhibition of CO₂-responsive neurons.

To distinguish between these possibilities, we tested the behaviour of *Or83b*² mutant flies in which most ORNs are non-functional, but electrophysiological responses to CO₂ are not affected¹⁴ (Fig. 1b). Consistent with the electrophysiological analysis, flies lacking *Or83b* have a robust avoidance response to CO₂, which is absent when 2,3-butanedione is included with CO₂ or is presented alone (Fig. 3b). Similar results, albeit with weaker effects, are obtained using 1-hexanol (Supplementary Fig. 5). Taken together, these results show that inhibitory odorants can effectively block CO₂-mediated innate avoidance behaviour.

CO₂ is one of the main components of dSO, which is emitted by stressed flies, and which triggers a robust avoidance behaviour in naive flies¹. We therefore examined whether 2,3-butanedione can disrupt avoidance to dSO. As reported previously¹, we found that naive flies avoid odour collected from a tube of vortexed flies (dSO), but not that collected from a tube of untreated flies (mock), in a T-maze assay. Remarkably, addition of 2,3-butanedione to dSO effectively abolishes avoidance behaviour (Fig. 3c).

Interestingly, we observed that with increasing concentrations of 2,3-butanedione, the CO₂ neuron is silenced well beyond the period of application (Fig. 3d). This effect is specific to 2,3-butanedione and is not observed for 1-hexanol (data not shown). To investigate this further, we exposed the fly to a 3-s stimulus of 2,3-butanedione (10⁻¹ dilution) and subsequently tested for the recovery of ab1C neuron responsiveness by applying a 0.5-s stimulus of 0.3% CO₂ every 30 s, over a period of 10 min (Fig. 3e). Surprisingly, the inhibitory effect of the initial exposure to 2,3-butanedione persisted for an extended period.

We wanted to test whether behaviour was also affected in a similar manner. We exposed flies for 1 min to 2,3-butanedione and then transferred them to clean air for 2 min before testing for CO₂-mediated avoidance behaviour. Remarkably, CO₂ avoidance is almost abolished in pre-treated flies (Fig. 3f). Prior exposure to another odorant 2-methyl phenol, which does not inhibit the CO₂ response (data not shown), does not have any effect on behaviour (Fig. 3f). Moreover, pre-exposure to 2,3-butanedione does not have a significant effect on behavioural attraction towards a different odorant, ethyl acetate (Fig. 3f). Taken together, these observations show that exposure to a long-term CO₂ response inhibitor can exert a profound and specific effect on the behaviour of the animal, even after it is no longer present in the environment. Similar observations were made with *Or83b* mutant flies (Fig. 3g).

To demonstrate unambiguously that 2,3-butanedione causes behaviour modification primarily by inhibiting CO₂ responsiveness of ab1C neurons and not by other peripheral or central mechanisms, we did the following experiment. We activated the ab1C neuron in a manner that is not inhibited by 2,3-butanedione, and asked whether 2,3-butanedione inhibits avoidance behaviour in this context. We

identified an odorant, butanone, which activates ab1C neurons (Supplementary Fig. 4a) strongly at 10⁻¹ dilution (Supplementary Fig. 6a) in a *Gr63a*-dependent manner (Supplementary Fig. 6b). We found that *Or83b* mutant flies strongly avoid butanone (10⁻¹ dilution) whereas flies lacking both *Or83b* and *Gr63a* do not (Fig. 3h), as predicted from the electrophysiology data. However, electrophysiological response to butanone is not affected by pre-exposure to, or the presence of, 2,3-butanedione (Supplementary Fig. 6c, d), unlike what we observed for CO₂. In a T-maze behaviour assay, 2,3-butanedione has no effect on behavioural avoidance of *Or83b* mutant flies to butanone, regardless of whether it is used to pre-treat the flies as described above or is included in a mixture with butanone (Fig. 3h). These results demonstrate that 2,3-butanedione disrupts CO₂ avoidance behaviour by directly inhibiting the CO₂ responsiveness of ab1C neurons, rather than by other indirect mechanisms.

CO₂ emitted in human breath is a critical component of odour blends used as host-seeking cues by many vector insect species that carry deadly diseases^{24–26}, including *Culex quinquefasciatus* mosquitoes that transmit filarial parasites in tropical countries, and West Nile virus in the USA and various parts of the world. *Culex* mosquitoes have three conserved proteins that are closely related to the *Drosophila* CO₂ receptors (data not shown), Gr21a and Gr63a (ref. 27). To test whether odorants that inhibit *Drosophila* CO₂ receptors can also inhibit CO₂ response in *Culex*, we tested CO₂-sensitive A neurons in peg sensilla on the surface of the maxillary palps of *Culex* mosquitoes using a panel of structurally related odours (Fig. 4a). We found that electrophysiological response to CO₂ is not inhibited by 2,3-butanedione, but is strongly inhibited by 1-butanal and 1-hexanol (Fig. 4b–d). These odours are the first reported inhibitors of CO₂-sensitive neurons in mosquitoes and may provide a valuable resource for the identification of economical, environmentally safe, volatile compounds that may reduce mosquito–human contact by blocking responsiveness to CO₂.

METHODS SUMMARY

Behavioural tests. T-maze behavioural testing using *Drosophila* stress odour, CO₂ and mixtures were performed as described¹, with some modifications (see Methods). The avoidance response was calculated as a preference index = (number of flies in test arm – number of flies in control arm)/(total number of flies in assay).

Electrophysiology. Extracellular single-unit recordings were performed as described previously²⁸ with some modifications (see Methods and Supplementary Information).

Genetics. Fly stocks were maintained on standard cornmeal medium at 25 °C. Wild-type stock was w¹¹¹⁸ backcrossed five generations to Canton S. The *Or83b*² mutant was obtained from the Bloomington stock centre. Stocks for *Δhalo*; *Or22a-Gal4* and *UAS-Gr21a* and *UAS-Gr63a* were gifts from J. Carlson. Additional lines of *Or22a-Gal4* were generated by mobilizing the original P-element insertion line using standard genetic techniques. The *Δhalo*; *Or22a-Gal4/UAS-Gr21a*, *UAS-Gr63a* flies were raised on standard cornmeal medium at 28 °C.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 April; accepted 17 July 2009.

Published online 26 August 2009.

1. Suh, G. S. B. *et al.* A single population of olfactory sensory neurons mediates an innate avoidance behaviour in *Drosophila*. *Nature* **431**, 854–859 (2004).
2. Suh, G. S. *et al.* Light activation of an innate olfactory avoidance response in *Drosophila*. *Curr. Biol.* **17**, 905–908 (2007).
3. Jones, W. D., Cayirlioglu, P., Kadow, I. G. & Vosshall, L. B. Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila*. *Nature* **445**, 86–90 (2007).
4. Guerenstein, P. G. & Hildebrand, J. G. Roles and effects of environmental carbon dioxide in insect life. *Annu. Rev. Entomol.* **53**, 161–178 (2008).
5. Hallem, E. A. & Sternberg, P. W. Acute carbon dioxide avoidance in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **105**, 8038–8043 (2008).
6. Clyne, P., Warr, C. & Carlson, J. Candidate taste receptors in *Drosophila*. *Science* **287**, 1830–1834 (2000).
7. Scott, K. *et al.* A chemosensory gene family encoding candidate gustatory and olfactory receptors in *Drosophila*. *Cell* **104**, 661–673 (2001).

8. Kwon, J. Y., Dahanukar, A., Weiss, L. A. & Carlson, J. R. The molecular basis of CO₂ reception in *Drosophila*. *Proc. Natl Acad. Sci. USA* **104**, 3574–3578 (2007).
9. Faucher, C., Forstreuter, M., Hilker, M. & de Bruyne, M. Behavioral responses of *Drosophila* to biogenic levels of carbon dioxide depend on life-stage, sex and olfactory context. *J. Exp. Biol.* **209**, 2739–2748 (2006).
10. Golding, J. B., Shearer, D., Wyllie, S. G. & McGlasson, W. B. Application of 1-MCP and propylene to identify ethylene-dependent ripening processes in mature banana fruit. *Postharvest Biol. Technol.* **14**, 87–98 (1998).
11. Golding, J. B., Shearer, D., McGlasson, W. B. & Wyllie, S. G. Relationships between respiration, ethylene, and aroma production in ripening banana. *J. Agric. Food Chem.* **47**, 1646–1651 (1999).
12. Fischler, W., Kong, P., Marella, S. & Scott, K. The detection of carbonation by the *Drosophila* gustatory system. *Nature* **448**, 1054–1057 (2007).
13. Katz, D. B. et al. Receptors, circuits, and behaviors: new directions in chemical senses. *J. Neurosci.* **28**, 11802–11805 (2008).
14. Larsson, M. C. et al. Or83b encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron* **43**, 703–714 (2004).
15. Benton, R., Sachse, S., Michnick, S. W. & Vosshall, L. B. Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors *in vivo*. *PLoS Biol.* **4**, e20 (2006).
16. Mayr, D., Mark, T., Lindinger, W., Brevard, H. & Yeretzian, C. Breath-by-breath analysis of banana aroma by proton transfer reaction mass spectrometry. *Int. J. Mass Spectrom.* **223–224**, 743–756 (2003).
17. Galliard, T., Matthew, J. A., Wright, A. J. & Fishwick, M. J. Enzymic breakdown of lipids to volatile and nonvolatile carbonyl fragments in disrupted tomato fruits. *J. Sci. Food Agric.* **28**, 863–868 (1977).
18. Martineau, B., Henick-Kling, T. & Acree, T. Reassessment of the influence of malolactic fermentation on the concentration of diacetyl in wines. *Am. J. Enol. Vitic.* **46**, 385–388 (1995).
19. Nykanen, L. & Nykanen, I. in *Volatile Compounds in Foods and Beverages* (ed. Maarse, H.) 548–580 (Dekker, 1991).
20. Hughes, P. S. & Baxter, E. D. in *Beer: Quality, Safety and Nutritional Aspects* 40–73 (Royal Society of Chemistry, 2001).
21. Carde, R. T. Odour plumes and odour-mediated flight in insects. *Ciba Found. Symp.* **200**, 54–66; discussion 66–70 (1996).
22. de Bruyne, M., Foster, K. & Carlson, J. Odor coding in the *Drosophila* antenna. *Neuron* **30**, 537–552 (2001).
23. Hallem, E. A. & Carlson, J. R. Coding of odors by a receptor repertoire. *Cell* **125**, 143–160 (2006).
24. Lu, T. et al. Odor coding in the maxillary palp of the malaria vector mosquito *Anopheles gambiae*. *Curr. Biol.* **17**, 1533–1544 (2007).
25. Zwiebel, L. J. & Takken, W. Olfactory regulation of mosquito–host interactions. *Insect Biochem. Mol. Biol.* **34**, 645–652 (2004).
26. Cooperband, M. F. & Carde, R. T. Orientation of *Culex* mosquitoes to carbon dioxide-baited traps: flight manoeuvres and trapping efficiency. *Med. Vet. Entomol.* **20**, 11–26 (2006).
27. Robertson, H. M. & Kent, L. B. Evolution of the gene lineage encoding the carbon dioxide receptor in insects. *J. Insect Sci.* **19**, 1–14 (2009).
28. Dobritsa, A. A., van der Goes van Naters, W., Warr, C. G., Steinbrecht, R. A. & Carlson, J. R. Integrating the molecular and cellular basis of odor coding in the *Drosophila* antenna. *Neuron* **37**, 827–841 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank W. van der Goes van Naters for help with setting up our electrophysiology rig, P. Atkinson and A. Kahlon for providing mosquitoes, J. Millar for providing reagents, and A. Dahanukar for commenting on the manuscript.

Author Contributions S.L.T. planned and performed the experiments and co-wrote the manuscript. A.R. planned the experiments, supervised the project and co-wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.R. (anand.ray@ucr.edu).

METHODS

Behavioural tests. T-maze behavioural testing using *Drosophila* stress odour, CO₂ and mixtures was performed as described¹, with the following modifications. The entire headspace from 15-ml capped 'emitter' or 'mock' fly tubes was withdrawn using fresh syringes and needles and infused into fresh capped 15-ml plastic tubes immediately before use in the T-maze. To test the response to mixtures, 10 µl of odorant diluted in paraffin oil (at the concentrations indicated) was placed on a Whatman filter paper (6-mm diameter) and placed carefully at the bottom of a fresh 15-ml plastic tube and capped about 10 min before starting the assay. The additional component (0.1 ml pure CO₂ or 15 ml dSO) was injected directly into this capped tube using a syringe, which was then used as the test arm in the T-maze. The tube in the control arm contained filter paper with 10 µl of paraffin oil solvent. The avoidance response was calculated as a preference index = (number of flies in test arm – number of flies in control arm)/(total number of flies in assay). Behavioural responses to CO₂ were tested using the T-maze by injecting 0.1 ml of pure CO₂ into a capped 15-ml tube with a syringe and needle immediately before the choice assay. For over-ripe fruits, fruits were allowed to ripen and ferment in a sealed plastic container for about 3 weeks, at which point 5 g of fruit paste was transferred to a fresh 50-ml plastic tube and sealed. After 5 min at room temperature, 15 ml of headspace was removed using a syringe, and transferred to a fresh 15-ml plastic tube that was used directly as the test arm of the T-maze. Yeast (1 g) was used to make a paste with 1 ml of 15% sucrose solution, and incubated at room temperature for 1 h in a 50-ml sealed tube. The cap was removed to release volatiles and then replaced; 15 ml of headspace was collected 5 min later and tested as described above. Similarly, 5-min collections of headspace were taken from 5 g of green fruits and 5 ml of beer (Stone Pale Ale, Stone Brewing Company). Before being tested for responses to headspace from fruit, beer and yeast, flies were pre-exposed to the same odours in separate 15-ml tubes for 2 min. To test the response to CO₂ or

other odours after prior exposure to odorants, 10 µl of odorant diluted in paraffin oil (10⁻² dilution unless otherwise indicated in legend) was loaded on a Whatman filter disc (6-mm diameter), which was placed carefully at the bottom of a fresh 15-ml plastic tube about 10 min before starting the assay. A small piece of cotton wool was inserted into the tube such that the flies were unable to make physical contact with the odorant-laden filter paper. Starved flies (24 h) were carefully put in the tube for 1 min and then transferred to a fresh tube containing room air for nearly 2 min. Just before the 2-min mark, the flies were transferred to the T-maze, and 0.1 ml of pure CO₂ was injected into one arm. The assay was started precisely at the 2 min mark and performed as usual for 1 min in the dark.

Electrophysiology. Extracellular single-unit recordings were performed as described previously²⁸. Odorant stimuli were delivered by Pasteur pipette odour cartridges as described previously²⁸ with some modifications (Supplementary Fig. 2). Chemicals were of the highest purity available, typically greater than 99% (Sigma-Aldrich). All odorants were diluted in paraffin oil. A controlled volume of air 5 ml s⁻¹ was puffed through the odour cartridge containing vapours, and was delivered into a constant humidified airstream of 10 ml s⁻¹ that was passed over the fly's antenna. The odorant vapour present in the cartridge was thus diluted approximately threefold, and the concentration of inhibitory odorants in the airstream that passed over the fly was significantly lower than that applied to the cartridge. CO₂ stimulus was pulsed through a separate delivery system that delivered controlled pulses (variable 2.5–6.5 ml s⁻¹) into the same humidified airstream, from either a 1%, 5% or 100% tank of CO₂ (Airgas). For delivery of binary mixtures of CO₂ with another odorant, we ensured a steady concentration of CO₂ to the fly preparation (Supplementary Fig. 2). Unless mentioned, responses were quantified by subtraction of spontaneous activity from activity during the stimulus. For each inhibitory odorant (ones that had a long-term effect on CO₂ response), each recording was obtained from a distinct unexposed fly, except in experiments in which only baseline activity was examined.

LETTERS

Nucleotides released by apoptotic cells act as a find-me signal to promote phagocytic clearance

Michael R. Elliott^{1,2}, Faraaz B. Cheken^{1,3}, Paul C. Trampont^{1,2}, Eduardo R. Lazarowski⁷, Alexandra Kadl⁴, Scott F. Walk^{1,2}, Daeho Park^{1,2}, Robin I. Woodson⁵, Marina Ostankovich⁴, Poonam Sharma⁴, Jeffrey J. Lysiak⁵, T. Kendall Harden⁸, Norbert Leitinger^{3,4} & Kodi S. Ravichandran^{1,2,6}

Phagocytic removal of apoptotic cells occurs efficiently *in vivo* such that even in tissues with significant apoptosis, very few apoptotic cells are detectable¹. This is thought to be due to the release of 'find-me' signals by apoptotic cells that recruit motile phagocytes such as monocytes, macrophages and dendritic cells, leading to the prompt clearance of the dying cells². However, the identity and *in vivo* relevance of such find-me signals are not well understood. Here, through several lines of evidence, we identify extracellular nucleotides as a critical apoptotic cell find-me signal. We demonstrate the caspase-dependent release of ATP and UTP (in equimolar quantities) during the early stages of apoptosis by primary thymocytes and cell lines. Purified nucleotides at these concentrations were sufficient to induce monocyte recruitment comparable to that of apoptotic cell supernatants. Enzymatic removal of ATP and UTP (by apyrase or the expression of ectopic CD39) abrogated the ability of apoptotic cell supernatants to recruit monocytes *in vitro* and *in vivo*. We then identified the ATP/UTP receptor P2Y₂ as a critical sensor of nucleotides released by apoptotic cells using RNA interference-mediated depletion studies in monocytes, and macrophages from P2Y₂-null mice³. The relevance of nucleotides in apoptotic cell clearance *in vivo* was revealed by two approaches. First, in a murine air-pouch model, apoptotic cell supernatants induced a threefold greater recruitment of monocytes and macrophages than supernatants from healthy cells did; this recruitment was abolished by depletion of nucleotides and was significantly decreased in P2Y₂^{-/-} (also known as P2ry2^{-/-}) mice. Second, clearance of apoptotic thymocytes was significantly impaired by either depletion of nucleotides or interference with P2Y₂ receptor function (by pharmacological inhibition or in P2Y₂^{-/-} mice). These results identify nucleotides as a critical find-me cue released by apoptotic cells to promote P2Y₂-dependent recruitment of phagocytes, and provide evidence for a clear relationship between a find-me signal and efficient corpse clearance *in vivo*.

Most developing thymocytes (95%) undergo apoptosis, but in the steady state only 1–2% are detectable as apoptotic^{4,5}. It has been speculated that dying thymocytes secrete soluble factors that attract resident phagocytes to promote prompt clearance^{2,6}. To determine whether apoptotic thymocytes release such factors, we assessed cell-free supernatants after apoptosis induction (by anti-Fas/CD95 cross-linking) for their ability to attract THP-1 monocytes or primary human monocytes in a transwell migration assay (Fig. 1a and Supplementary Fig. 2). Apoptotic supernatants caused a threefold increase in monocyte migration than supernatants from live thymocytes did. This release of chemotactic factors was also seen with Jurkat cells (a mature T-cell line) induced to undergo Fas-mediated or

ultraviolet-mediated apoptosis (Fig. 1b). There was no detectable increase in membrane permeability or leakage of cytoplasmic markers when the supernatants were collected (Fig. 2g and Supplementary Fig. 1a–e). Moreover, supernatants from cells pretreated with the caspase inhibitor zVAD-fmk before the induction of apoptosis failed to induce monocyte migration (Fig. 1a, c), suggesting caspase-dependent and regulated release of chemoattractant(s). In a time course, release of chemotactic factor(s) was correlated with the onset and progression of apoptosis (assessed by annexin V binding to phosphatidylserine exposed on apoptotic cells, and caspase-3/7 activity; Fig. 1b and Supplementary Fig. 1b, c). Last, the chemotactic factor(s) were soluble and heat-stable, because high-speed centrifugation or boiling of the supernatants did not affect the chemotactic potential (Supplementary Fig. 3f).

We next tested whether find-me signal(s) in apoptotic cell supernatants could attract phagocytes *in vivo*. We used a murine dorsal air-pouch model (Fig. 1d) in which the supernatants from apoptotic or live cells were injected into sterile, subcutaneous air-pouches⁷. When cells in the air-pouch were recovered by lavage after 24 h, apoptotic cell supernatants caused a threefold increase in the number of CD45⁺ leukocytes recruited in comparison with live cell supernatants or medium alone (Fig. 1e; $n = 8$, $P = 0.02$). The total number of monocytes and macrophages (CD11b⁺/Gr-1^{low}) in the lavage was increased about threefold in comparison with neutrophils (Gr-1^{high} cells) (Fig. 1e). By contrast, injection of bacterial lipopolysaccharide induced the recruitment of mostly Gr-1^{high} neutrophils (Fig. 1f). This is consistent with previous studies on the preferential recruitment of monocyte/macrophages rather than inflammatory neutrophils to cells undergoing apoptosis^{8,9}. F4/80⁺ macrophages recruited to the pouch could also engulf apoptotic Jurkat cells injected into the pouch (data not shown). These data revealed the release of find-me signal(s) by apoptotic lymphocytes that attract monocytes *in vitro* and *in vivo*.

We then sought to determine the nature of the chemoattractant. On the basis of *in vitro* studies, the lipid lysophosphatidylcholine was implicated as a find-me signal released by apoptotic MCF-7 cancer cells¹⁰. However, we did not observe monocyte migration towards purified lysophosphatidylcholine over a range of concentrations (0.1–100 μ M; data not shown) under our conditions; moreover, treatment of apoptotic cell supernatants with phospholipase D, an enzyme that promotes the hydrolysis of lysophosphatidylcholine (see Supplementary Fig. 3a, b), did not affect chemotactic activity of supernatants from apoptotic thymocytes, Jurkat cells or MCF-7 cells (Fig. 1a, j, k). CX₃CL1 (fractalkine) released by apoptotic Burkitt lymphoma B cells can also act as a find-me signal¹¹; however, THP-1

¹Berne B. Carter Center for Immunology Research, ²Center for Cell Clearance, ³Department of Pharmacology, ⁴Robert M. Berne Cardiovascular Research Center, ⁵Department of Urology, ⁶Department of Microbiology, University of Virginia, Charlottesville, Virginia 22908, USA. ⁷Department of Medicine, ⁸Department of Pharmacology, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599, USA.

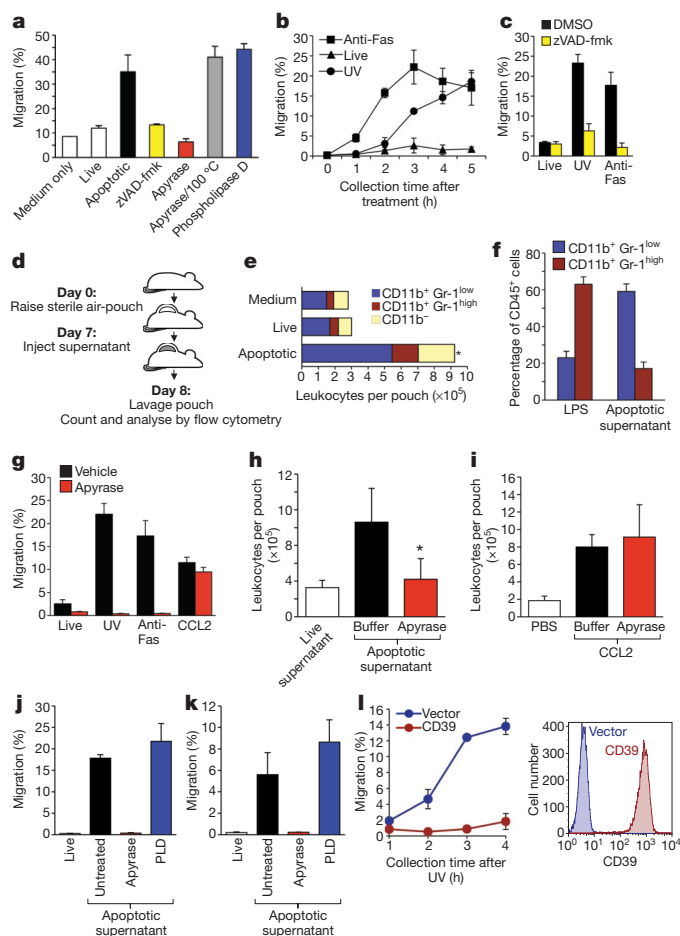


Figure 1 | Chemotactic factor released by apoptotic cells attracts monocytes *in vitro* and *in vivo*. **a**, Migration of THP-1 monocytes through a transwell (5-μm pore size) to supernatants from control ('live') or Fas-induced apoptotic murine thymocytes, thymocytes pretreated with caspase inhibitor (zVAD-fmk), and apoptotic cell supernatants with apyrase, heat-inactivated apyrase or phospholipase D. The fraction of input monocytes that migrated to the lower chamber is shown. Fas treatment was for 1 h. **b**, Attraction of monocytes by Jurkat T-cell supernatants collected at the indicated times after apoptosis induction by means of ultraviolet (UV) or anti-Fas for the indicated durations. **c**, Monocyte attraction was inhibited by pretreatment of Jurkat cells with zVAD-fmk before ultraviolet or anti-Fas treatment. DMSO, dimethylsulphoxide. **d**, Schematic diagram for testing the recruitment of leukocytes by apoptotic cell supernatants in the mouse air-pouch model. **e**, Recruitment to the air-pouch of macrophages and monocytes (CD11b⁺/Gr-1^{low}) or neutrophils (CD11b⁺/Gr-1^{high}) 24 h after injection of apoptotic cell supernatants. There were eight mice per group. Asterisk, $P = 0.02$. **f**, Monocyte/macrophage and neutrophil populations recruited to the air-pouch 24 h after injection of lipopolysaccharide (1 μg) or apoptotic supernatants. Results are the average of six (lipopolysaccharide) or nine (apoptotic supernatant) mice. **g–i**, Treatment of apoptotic cell supernatants with apyrase inhibits attraction of monocytes *in vitro* (g) or in the air-pouch model *in vivo* (h, i), but does not affect monocyte migration towards the chemokine CCL2 (250 ng ml⁻¹ *in vivo* or 50 ng ml⁻¹ *in vitro*). There were five (h) or three (i) mice per group. Asterisk, $P = 0.005$. **j, k**, Migration of monocytes towards supernatants from apoptotic Jurkat (j) or MCF-7/caspase-3 (k) cells, supernatants being treated with apyrase or phospholipase D (PLD). **l**, Right: CD39 surface expression on transfected Jurkat cells. Left: monocyte migration towards supernatants from CD39-overexpressing cells after treatment with ultraviolet. Error bars indicate s.e.m. throughout.

monocytes used here failed to show migration towards purified CX₃CL1, and the anti-fractalkine-depleting antibody did not block migration in our assays (data not shown). Thus, the find-me signal released by apoptotic primary thymocytes and Jurkat cells seemed to be distinct from those reported previously.

Subsequently, several lines of evidence suggested a role for extracellular nucleotides as a possible find-me signal. Treatment of apoptotic cell supernatants with recombinant apyrase, an enzyme that hydrolyses nucleoside triphosphates and diphosphates to nucleoside monophosphates (for example ATP→ADP→AMP), abolished the monocyte chemoattractant activity of apoptotic thymocytes, Jurkat and MCF-7 cells at all time points (Fig. 1a, g, j, k and Supplementary Fig. 3e). Apyrase did not affect monocyte migration towards chemokines CCL2 or CXCL12 (Fig. 1g and data not shown). Treatment of apoptotic cell supernatants (but not CCL2) with apyrase before injection into the dorsal air-pouch also inhibited the attraction of leukocytes *in vivo* (Fig. 1h, i). As another approach, we expressed in Jurkat cells the transmembrane protein CD39 (NTPDase-1), the primary mammalian ecto-apyrase responsible for NTP degradation by immune cells *in vivo*¹² (see Supplementary Fig. 4b); CD39 expression abrogated the chemoattractant activity in the supernatants of apoptotic Jurkat cells (Fig. 1l). Neither treatment with apyrase nor CD39 overexpression impaired the induction of apoptosis (see below and Supplementary Fig. 4a). Inactivation of apyrase by heat (before addition to apoptotic cell supernatants) abolished its effect (Fig. 1a and Supplementary Fig. 3d), suggesting a need for intact enzymatic activity. Thus, induction of apoptosis led to an accumulation of extracellular nucleotides capable of monocyte chemoattraction *in vitro* and *in vivo*.

Among the four naturally occurring extracellular nucleotides (ATP, ADP, UTP and UDP), ATP and UTP induced strong chemotactic activity in THP-1 monocytes (Fig. 2a); in contrast, ADP and UDP showed partial activity at the highest concentrations tested (Fig. 2a), but lower than that of NTPs. The migration was also stimulated by the non-hydrolysable ATP analogue ATP-γS but not by adenosine (Supplementary Fig. 5a), suggesting attraction primarily towards triphosphate nucleotides. When ATP and UTP levels in apoptotic cell supernatants were directly quantified (see Methods), higher ATP and UTP levels could be detected as early as 2 h after the induction of apoptosis, with a further increase by 4 h (Fig. 2b). The concentrations of ATP and UTP at the time point when the apoptotic supernatants induced maximal monocyte migration correlated well with the concentrations at which pure ATP and UTP caused maximal migration (about 100 nM) (Figs 1b and 2a, b). The addition of pure ATP and UTP to the upper chamber of the transwell, to disrupt the gradient, blocked the migration of monocytes to the lower chamber containing apoptotic cell supernatants (Supplementary Fig. 6a–c). Although ATP can promote chemokinesis or random migration of neutrophils¹³, the addition of pure ATP or UTP only to the upper chamber did not induce the migration of THP-1 cells to the lower chamber, indicating that the migration induced was not chemokinesis but chemotaxis (Supplementary Fig. 6d). Furthermore, the release of nucleotides from apoptotic cells is dependent on caspase activity, and occurs after different types of apoptosis induction (DNA damage, receptor-mediated and steroid-induced) and occurred in primary thymocytes, Jurkat and epithelial cells undergoing apoptosis; this release of ATP was also well correlated with the induction of apoptosis (Fig. 2c–e and Supplementary Figs 1b–d and 8). The release of ATP during apoptosis was not due to leakage of cellular contents or to mechanical stress during handling of cells¹⁴ (Fig. 2f, g). Collectively, these data strongly suggested a role for ATP and UTP as find-me signals that are important for phagocyte chemoattraction by apoptotic cells.

We then addressed how the phagocytes might 'sense' the extracellular nucleotides as a find-me cue. Leukocyte migration towards nucleotides has been shown to be dependent on members of the P2Y family of G-protein-coupled receptors^{15,16}. We tested the role of P2Y receptors on monocytes and macrophages in migration towards apoptotic cell supernatants. First, pretreatment of THP-1 cells with suramin, a non-selective inhibitor of P2 family members, showed dose-dependent inhibition of migration towards apoptotic supernatants but not towards CCL2 (Fig. 3a and Supplementary Fig. 5c). After several P2Y family members were evaluated through known antagonists (scored by inhibition of migration towards

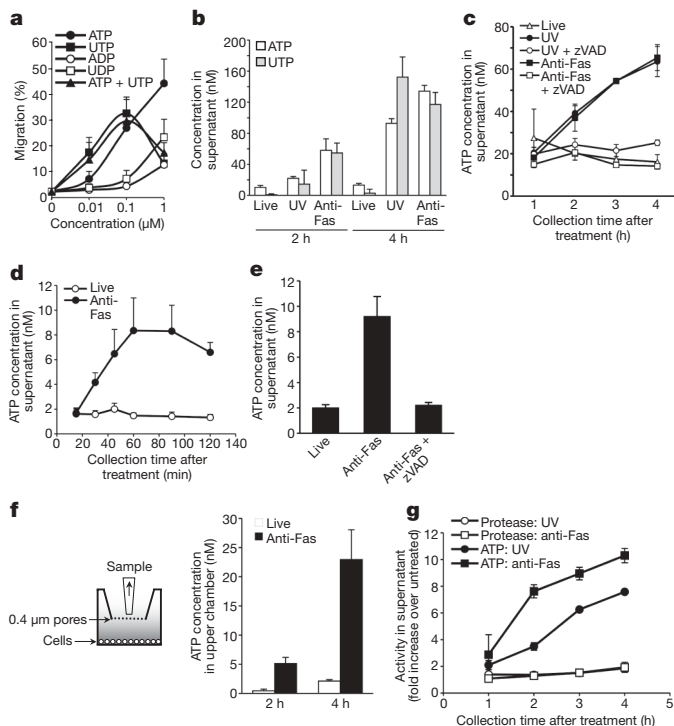


Figure 2 | Regulated release of ATP and UTP as chemoattractants by apoptotic cells. **a**, The percentage of migrated THP-1 monocytes towards purified nucleotides at the indicated concentrations. **b**, Quantification of ATP and UTP in supernatants of control and apoptotic Jurkat cells at 2 and 4 h after induction of apoptosis. **c**, ATP level in supernatants of apoptotic Jurkat cells at the indicated times, and inhibition by zVAD-fmk (zVAD). **d**, ATP levels in supernatants of control or Anti-Fas-treated thymocytes for the indicated durations. **e**, ATP levels in supernatants of thymocytes treated with zVAD-fmk before Anti-Fas treatment (1 h). **f**, Left: diagram of supernatant collection without disturbing the cells. Right: quantification of ATP that has diffused through the 0.4- μ m filter into the medium from untreated (live) or Anti-Fas-treated Jurkat cells. **g**, The integrity of the cell membrane is retained when apoptotic cell supernatants are collected, as determined by ATP release but not leakage of cytoplasmic protease activity. Data are presented as the fold change in ATP-dependent or protease-dependent luminescence in the supernatants of Jurkat cells relative to untreated cells at the indicated times after treatment with Anti-Fas or ultraviolet. Error bars indicate s.e.m. throughout.

apoptotic cell supernatants), we focused on $P2Y_2$ based on its known affinities for ATP and UTP (because both are released by apoptotic cells), and the $P2Y_2$ expression on monocytes and macrophages¹⁶. Although $P2Y_4$ also fits the profile for ATP and UTP binding, THP-1 cells express no detectable $P2Y_4$ but strongly express $P2Y_2$ mRNA (ref. 17 and data not shown). Short interfering RNA (siRNA)-mediated knockdown of $P2Y_2$ in THP-1 monocytes led to a 60% decrease in $P2Y_2$ mRNA and also partly inhibited migration towards apoptotic cell supernatants ($10.5 \pm 0.8\%$ versus $16.3 \pm 2.0\%$ (means \pm s.e.m.) for control siRNA; $n = 6$, $P = 0.003$), but did not affect migration towards CCL2 or CXCL12 (Fig. 3b and data not shown). Bone-marrow-derived macrophages (BMDMs) from $P2Y_2^{-/-}$ mice³ showed impaired migration towards apoptotic supernatants, but their migration towards CXCL12 was intact (Fig. 3c). When apoptotic cell supernatants were injected into the air-pouch of $P2Y_2$ -deficient mice, there was a strong decrease in the recruitment of monocytes and macrophages to the pouch, indicating *in vivo* relevance of this receptor in sensing the find-me signal (Fig. 3d). Although $P2Y_6$ was shown to have a function in UDP-dependent leukocyte migration^{15,18} and phagocytosis by microglial cells¹⁹, neither RNA-mediated interference towards $P2Y_6$ nor towards the $P2Y_6$ antagonist MRS2578 showed specific inhibition of migration (Supplementary Fig. 5b and data not shown). Finally, addition of antagonists to the adenosine

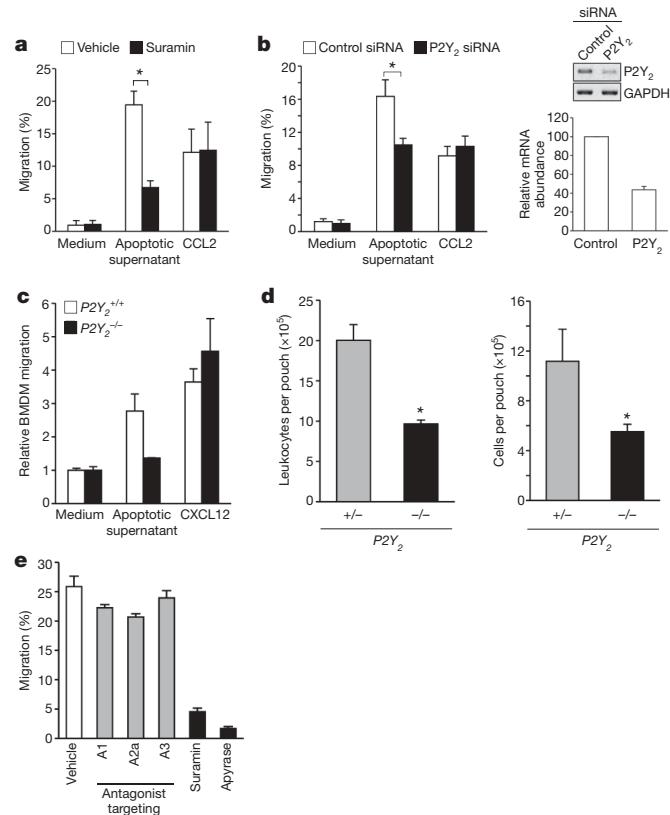


Figure 3 | $P2Y_2$ receptor on monocytes and macrophages as a sensor of ATP and UTP released by apoptotic cells. **a**, Effect of pretreatment of monocytes with $P2Y_2$ receptor antagonist Suramin ($100 \mu\text{M}$) on migration towards apoptotic cell supernatants or the chemokine CCL2 (50 ng ml^{-1}). Asterisk, $P = 0.003$; $n = 3$. **b**, Migration of THP-1 monocytes transfected with siRNA specific for $P2Y_2$ receptor or control siRNA. Asterisk, $P = 0.03$; $n = 6$. Right: real-time polymerase chain reaction and agarose-gel electrophoresis (inset, inverted image) analysis of $P2Y_2$ receptor mRNA levels in siRNA-transfected THP-1 cells. **c**, BMDMs from $P2Y_2^{+/+}$ or $P2Y_2^{-/-}$ mice were assessed for transwell migration towards apoptotic Jurkat supernatant or CXCL12 (50 ng ml^{-1}). **d**, Recruitment of CD45^+ cells (left) and $\text{CD11b}^+/\text{Gr-1}^{\text{low}}$ monocytes and macrophages (right) to the air-pouch of mice with the indicated $P2Y_2$ genotypes 24 h after injection of apoptotic Jurkat supernatants. There were five ($P2Y_2^{+/+}$) or seven ($P2Y_2^{-/-}$) mice per group. Asterisk, $P \leq 0.03$. **e**, THP-1 cells pretreated with antagonists targeting adenosine receptors A1, A2a and A3, apyrase or Suramin before migration assay to apoptotic supernatants. Error bars indicate s.e.m. throughout.

receptors A1, A2a and A3 (refs 13, 20) or the A2 receptor agonist CGS21680 did not significantly affect the migration of monocytes to the apoptotic cell supernatants (Fig. 3e and Supplementary Fig. 5d). Moreover, adenosine itself did not induce migration of THP-1 cells, and addition of exogenous adenosine did not affect migration towards apoptotic cell supernatants (Supplementary Fig. 5a, d). Taken together, these results identify the $P2Y_2$ receptor on monocytes and macrophages as a critical sensor of the find-me signal released by apoptotic cells.

As a further test of the importance of nucleotides as a find-me signal in an *in vivo* model of apoptosis, we used intraperitoneal injection of dexamethasone, in which a large fraction of immature thymocytes undergo largely synchronous apoptosis and phagocytic clearance^{5,21}. Dexamethasone injection induced thymic apoptosis, with a decline in thymus size and cellularity within 4 h, and a decrease to less than half that of control-treated mice by 8 h (Fig. 4a). Treatment of thymocytes with dexamethasone *in vitro* also induced apoptosis in a large fraction of the cells by 4 h (more than 40%) and 6 h (more than 60%) (Fig. 4g). We examined whether apyrase-mediated destruction of nucleotides *in vivo* could influence the

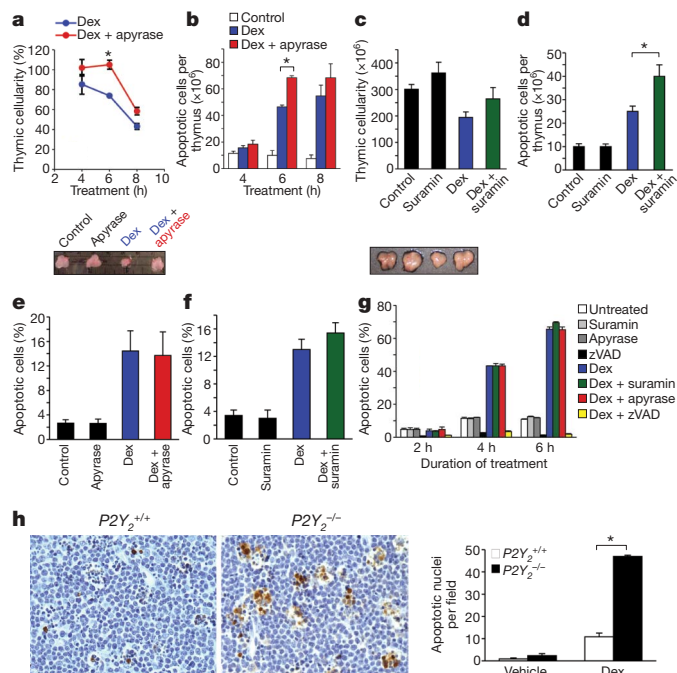


Figure 4 | Interference with the nucleotide find-me signal or its sensing impairs the clearance of apoptotic cells in the thymus. **a, b**, C57BL/6 mice (4–6 weeks old) were injected intraperitoneally with 250 μ g of dexamethasone (Dex) for the indicated durations, with or without apyrase, and the cellularity (expressed as a percentage of the control-treated animals within the same experimental group (4, 6 or 8 h)) (**a**) and number of apoptotic cells per thymus (**b**) were determined (annexin V-positive/propidium iodide-negative populations). A representative thymus from mice treated as indicated (6 h dexamethasone treatment) is shown below **a**. Data shown are representative of two to four experiments per time point, with at least three mice per group. Asterisk, $P = 0.03$. **c, d**, As in **a** and **b** except that mice were injected with 6 mg of suramin 1 h before dexamethasone injection (6 h). A representative thymus from each group is shown below **c**. There were four mice per group. Asterisk, $P = 0.03$. **e, f**, Effect of apyrase (**e**) and suramin (**f**) on the percentage of apoptotic cells *in vivo* in the thymuses of 6 h dexamethasone-treated mice. **g**, Apyrase (0.05 U ml⁻¹) and suramin (100 μ M) do not affect dexamethasone-induced thymocyte apoptosis *in vitro*. zVAD-fmk (zVAD) was included as a control. The percentage apoptotic cells is shown. **h**, Left: paraffin-embedded sections from thymuses of wild-type and P2Y₂^{-/-} mice treated with dexamethasone for 6 h (brown, Apostain; blue, haematoxylin). Right: mean number of Apostain-positive nuclei per field from 10–16 random fields per section per mouse. Asterisk, $P = 0.001$. Error bars indicate s.e.m. throughout.

apoptosis and clearance of dexamethasone-mediated thymocytes. Injection of apyrase significantly reversed the decline in thymus size and cellularity resulting from dexamethasone treatment (especially at 6 h) (Fig. 4a). This was not due to an effect of apyrase on the apoptotic process itself, because the fraction of cells undergoing apoptosis as a result of dexamethasone treatment was unchanged on treatment with apyrase *in vivo* or *in vitro* (Fig. 4e, g). The total numbers of apoptotic cells remaining at 6 and 8 h were greater in mice treated with apyrase plus dexamethasone than in those treated with dexamethasone alone (Fig. 4b). Because apyrase had no effect on the phagocytic capacity of macrophages (Supplementary Fig. 9), apyrase-mediated destruction of the nucleotide find-me signal seemed to affect phagocyte recruitment, and in turn to delay clearance.

In a complementary set of studies, we tested how a disruption to the 'sensing' of the find-me signal affected apoptotic cell clearance. Injection of the P2Y inhibitor suramin before dexamethasone reversed the diminution of the thymic cellularity and organ size seen with dexamethasone alone (Fig. 4c, d). The total number of apoptotic thymocytes also increased in suramin plus dexamethasone conditions (Fig. 4d). As with apyrase, suramin itself did not alter the

induction of thymocyte apoptosis (Fig. 4f, g). When we assessed the presence of apoptotic cells in the native thymic architecture (by immunohistochemistry with antibody against single-stranded DNA), there were greater numbers of uncleared apoptotic cells after treatment with suramin plus dexamethasone than after treatment with dexamethasone alone (Supplementary Fig. 10). We also assessed whether genetic disruption of the putative find-me signal receptor P2Y₂ would affect apoptotic cell clearance *in vivo*. After injection of dexamethasone, the number of apoptotic thymocytes in the thymuses of P2Y₂^{-/-} mice was significantly greater than in control mice (Fig. 4h). Taken together, these results show that the disruption of a find-me signal circuit at the level of nucleotides or the sensing receptor significantly impairs the clearance of apoptotic thymocytes, without an apparent effect on the induction of apoptosis or engulfment.

The data presented here provide new insights into particular aspects of the apoptotic cell death process. First, this work identifies ATP and UTP as a critical and non-redundant find-me signal released by apoptotic cells, documenting the regulated and caspase-dependent release of nucleotides from apoptotic cells with a functional secondary consequence. Because nucleotide release is seen in primary cells and cell lines (after different types of apoptotic stimulus), nucleotides may be a broadly used find-me signal. However, these data do not exclude the possibility of other chemotactic factors that work alone or together with nucleotides. Second, these data establish a clear relationship between a find-me signal and efficient apoptotic cell clearance *in vivo*; the disruption of the find-me signal circuit at the level of ATP/UTP or the receptors (P2Y) impaired the clearance of apoptotic thymocytes. Although we focused here on motile monocytes/macrophages, genetic studies in *Caenorhabditis elegans*, in which healthy cells engulf the dying neighbours, have revealed a link between apoptosis and engulfment^{22,23}. How nucleotides might regulate engulfment by neighbouring non-professional phagocytes^{6,24} remains to be determined.

Extracellular nucleotides at higher concentrations are considered pro-inflammatory²⁵ (more than 1 μ M, for example, by necrotic cells; Supplementary Fig. 7), but nucleotides can also induce an anti-inflammatory response²⁶. Besides serving as a find-me signal, it remains to be determined whether nucleotides participate in anti-inflammatory signalling during engulfment. It has recently been shown that lactoferrin released by apoptotic cells inhibits neutrophil migration²⁷. How nucleotides and lactoferrin concurrently promote monocyte migration while inhibiting neutrophil migration remains to be seen. Because a failure to clear dying cells promptly can lead to autoimmunity and chronic inflammatory diseases⁹, phagocyte chemoattraction to apoptotic cells by means of nucleotides may have implications for human disease states related to failed clearance.

METHODS SUMMARY

Supernatant preparation. Jurkat (E6-1) cells at 2×10^6 ml⁻¹ in RPMI medium containing 5% heat-inactivated FBS and 10 mM HEPES, pH 7.2 were treated with 250 ng ml⁻¹ anti-Fas (CH11) or 100 mJ cm⁻² of ultraviolet-C. Freshly isolated thymocytes from 4–5-week-old C57BL/6 mice were treated at 5×10^6 ml⁻¹ with crosslinked anti-Fas (5 μ g ml⁻¹ Jo2, 2 μ g ml⁻¹ protein G) in RPMI containing 1% BSA and 10 mM HEPES, pH 7.2. Supernatants were collected by two successive centrifugations at 500g for 4 min at 4 °C. Pretreatment with zVAD-fmk (50 μ M) and treatment of supernatants with apyrase (0.025 U ml⁻¹) and phospholipase D (0.5 U ml⁻¹) was performed for 5 min at room temperature (25 °C). MCF-7 supernatants were prepared as described previously¹⁰.

Migration assays. THP-1 cells at 2×10^6 ml⁻¹ were placed for 1 h on a 5- μ m pore-size Transwell with chemoattractant. Percentage migration was determined by flow cytometry with AccuCount beads. For migration of BMDMs, 5×10^4 d7 BMDM cells were placed on a 5- μ m pore-size Transwell and incubated for 2 h with Jurkat supernatants in RPMI containing 1% BSA and 10 mM HEPES, pH 7.2 in the lower chamber; migration was determined by Diff-Quick staining and microscopy.

In vivo experiments. Air-pouch experiments were performed as described⁷, with 8–12-week old C57BL/6 mice injected with 1 ml of 0.2- μ m-filtered supernatants. After 24 h, the pouch was lavaged and cells were counted and analysed by flow cytometry. For thymic clearance studies, 4–5-week-old C57BL/6 mice were injected intraperitoneally with 250 μ g dexamethasone and 5 U apyrase at the

midpoint of dexamethasone treatment, and with 6 mg suramin 1 h before dexamethasone injection. Thymocytes were stained with annexin V and propidium iodide and beads were added for quantification by flow cytometry. For *in situ* apoptotic cell detection, thymuses from female $P2Y_2^{-/-}$ or age-matched C57BL/6 mice were stained by Apostain as described²⁸.

Nucleotide measurement. ATP was measured by luciferase reaction as described²⁹. UTP was quantified by the UDP-glucose pyrophosphorylase-based reaction³⁰. Measurements were conducted on supernatants prepared from cells cultured in RPMI containing 1% BSA and 10 mM HEPES, pH 7.2.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 June; accepted 16 July 2009.

- Henson, P. M. & Hume, D. A. Apoptotic cell removal in development and tissue homeostasis. *Trends Immunol.* **27**, 244–250 (2006).
- Lauber, K., Blumenthal, S. G., Waibel, M. & Wesselborg, S. Clearance of apoptotic cells: getting rid of the corpses. *Mol. Cell* **14**, 277–287 (2004).
- Homolya, L., Watt, W. C., Lazarowski, E. R., Koller, B. H. & Boucher, R. C. Nucleotide-regulated calcium signaling in lung fibroblasts and epithelial cells from normal and $P2Y_2$ receptor ($-/-$) mice. *J. Biol. Chem.* **274**, 26454–26460 (1999).
- Hogquist, K. A., Baldwin, T. A. & Jameson, S. C. Central tolerance: learning self-control in the thymus. *Nature Rev. Immunol.* **5**, 772–782 (2005).
- Surh, C. D. & Sprent, J. T-cell apoptosis detected *in situ* during positive and negative selection in the thymus. *Nature* **372**, 100–103 (1994).
- Ravichandran, K. S. & Lorenz, U. Engulfment of apoptotic cells: signals for a good meal. *Nature Rev. Immunol.* **7**, 964–974 (2007).
- Kadl, A., Galkina, E. & Leitinger, N. Induction of CCR2-dependent macrophage accumulation by oxidized phospholipids in the air-pouch model of inflammation. *Arthritis Rheum.* **60**, 1362–1371 (2009).
- Huynh, M. L., Fadok, V. A. & Henson, P. M. Phosphatidylserine-dependent ingestion of apoptotic cells promotes TGF- β 1 secretion and the resolution of inflammation. *J. Clin. Invest.* **109**, 41–50 (2002).
- Savill, J., Dransfield, I., Gregory, C. & Haslett, C. A blast from the past: clearance of apoptotic cells regulates immune responses. *Nature Rev. Immunol.* **2**, 965–975 (2002).
- Lauber, K. *et al.* Apoptotic cells induce migration of phagocytes via caspase-3-mediated release of a lipid attraction signal. *Cell* **113**, 717–730 (2003).
- Truman, L. A. *et al.* CX3CL1/fractalkine is released from apoptotic lymphocytes to stimulate macrophage chemotaxis. *Blood* **112**, 5026–5036 (2008).
- Mizumoto, N. *et al.* CD39 is the dominant Langerhans cell-associated ecto-NTPDase: modulatory roles in inflammation and immune responsiveness. *Nature Med.* **8**, 358–365 (2002).
- Chen, Y. *et al.* ATP release guides neutrophil chemotaxis via $P2Y_2$ and A3 receptors. *Science* **314**, 1792–1795 (2006).
- Lazarowski, E. R., Homolya, L., Boucher, R. C. & Harden, T. K. Direct demonstration of mechanically induced release of cellular UTP and its implication for uridine nucleotide receptor activation. *J. Biol. Chem.* **272**, 24348–24354 (1997).
- Myrtek, D. & Idzko, M. Chemotactic activity of extracellular nucleotides on human immune cells. *Purinergic Signal.* **3**, 5–11 (2007).
- Burnstock, G. & Knight, G. E. Cellular distribution and functions of P2 receptor subtypes in different systems. *Int. Rev. Cytol.* **240**, 31–304 (2004).
- Moore, D. J. *et al.* Expression pattern of human $P2Y$ receptor subtypes: a quantitative reverse transcription-polymerase chain reaction study. *Biochim. Biophys. Acta* **1521**, 107–119 (2001).
- Idzko, M. *et al.* Characterization of the biological activities of uridine diphosphate in human dendritic cells: influence on chemotaxis and CXCL8 release. *J. Cell. Physiol.* **201**, 286–293 (2004).
- Koizumi, S. *et al.* UDP acting at $P2Y_6$ receptors is a mediator of microglial phagocytosis. *Nature* **446**, 1091–1095 (2007).
- Hasko, G., Linden, J., Cronstein, B. & Pacher, P. Adenosine receptors: therapeutic aspects for inflammatory and immune diseases. *Nature Rev. Drug Discov.* **7**, 759–770 (2008).
- Kawane, K. *et al.* Impaired thymic development in mouse embryos deficient in apoptotic DNA degradation. *Nature Immunol.* **4**, 138–144 (2003).
- Hoepfner, D. J., Hengartner, M. O. & Schnabel, R. Engulfment genes cooperate with ced-3 to promote cell death in *Caenorhabditis elegans*. *Nature* **412**, 202–206 (2001).
- Reddien, P. W., Cameron, S. & Horvitz, H. R. Phagocytosis promotes programmed cell death in *C. elegans*. *Nature* **412**, 198–202 (2001).
- Monks, J., Smith-Steinhart, C., Kruk, E. R., Fadok, V. A. & Henson, P. M. Epithelial cells remove apoptotic epithelial cells during post-lactation involution of the mouse mammary gland. *Biol. Reprod.* **78**, 586–594 (2008).
- Kono, H. & Rock, K. L. How dying cells alert the immune system to danger. *Nature Rev. Immunol.* **8**, 279–289 (2008).
- la Sala, A. *et al.* Alerting and tuning the immune response by extracellular nucleotides. *J. Leukoc. Biol.* **73**, 339–343 (2003).
- Bournazou, I. *et al.* Apoptotic human cells inhibit migration of granulocytes via release of lactoferrin. *J. Clin. Invest.* **119**, 20–32 (2009).
- Lysiak, J. J., Turner, S. D. & Turner, T. T. Molecular pathway of germ cell apoptosis following ischemia/reperfusion of the rat testis. *Biol. Reprod.* **63**, 1465–1472 (2000).
- Lazarowski, E. R., Boucher, R. C. & Harden, T. K. Constitutive release of ATP and evidence for major contribution of ecto-nucleotide pyrophosphatase and nucleoside diphosphokinase to extracellular nucleotide concentrations. *J. Biol. Chem.* **275**, 31061–31068 (2000).
- Lazarowski, E. R. & Harden, T. K. Quantitation of extracellular UTP using a sensitive enzymatic assay. *Br. J. Pharmacol.* **127**, 1272–1278 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Rock, C. Borowski and members of the Ravichandran laboratory for helpful suggestions; I. Juncadella for lung epithelial cells; K. Lauber and S. Wesselborg for providing MCF-7/caspase-3 cells; and R. Tacke for assistance with primary monocyte experiments. This work was supported by Public Health Service grants from the National Institutes of Health (to K.S.R. and N.L.), the American Cancer Society (to M.R.E.) and the University of Virginia Farrow Fellowship (to M.R.E.).

Author Contributions M.R.E. designed, performed and analysed most of the experiments in this study, with input from K.S.R. ATP quantification experiments were performed by F.B.C., and P.T.C. assisted with *in vivo* thymic apoptosis experiments. E.R.L. performed the high-performance liquid chromatography analysis of supernatants. S.F.W. generated the CD39 expression plasmid and stable Jurkat cell lines. D.P. conducted phagocytosis experiments. A.K. and N.L. performed the mass spectrometry analysis and provided critical support in establishing the air-pouch model system. R.I.W. and J.J.L. conducted immunohistochemical detection of apoptotic cells in the thymus. M.O. and P.S. assisted with the BMDM generation and macrophage chemotaxis experiments. T.K.H. provided critical intellectual input in the preparation of the manuscript. K.S.R. provided overall coordination with respect to conception, design and supervision of the study. K.S.R. and M.R.E. wrote the manuscript with comments from co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing interests. Correspondence and requests for materials should be addressed to K.S.R. (ravi@virginia.edu).

METHODS

Reagents. Purified nucleotides, adenosine, dexamethasone, etoposide, protein G, lipopolysaccharide and suramin were obtained from Sigma-Aldrich. Annexin V and other flow cytometry reagents were obtained from eBioscience. Other reagents were obtained as follows: recombinant apyrase (New England Biolabs), zVAD-fmk (Alexis Biochemicals), phospholipase D and ATP- γ S (EMD), anti-Fas (Jo2 anti-mouse clone, Becton Dickinson; CH11 anti-human clone, Millipore), siRNA (Dharmacon), purified lipids (Avanti Polar Lipids), chemokines (R&D). Adenosine receptor reagents were a gift from J. Linden; the antagonists used (20 μ M each) were 8-cyclopentyl-1,3-dipropylxanthine (A1), ZM241 (A2a) and MRS1191 (A3). Experiments were performed comparing the stock solution of apyrase with the same solution dialysed into PBS. Although the activity of the enzyme towards pure ATP *in vitro* was slightly decreased after dialysis of the enzyme (see Supplementary Fig. 3c), the two preparations performed the same in all migration experiments examined.

Air-pouch and thymic apoptotic cell clearance studies. All animal studies were performed in accordance with the University of Virginia Animal Care and Use Committee guidelines; animals were housed in a specific pathogen-free facility. For air-pouch experiments, female C57BL/6 mice were used (Charles River Laboratories), except those conducted with P2Y₂-deficient mice, which included males and females (results were similar). P2Y₂-deficient mice were obtained from Taconic³, with permission from B. Koller. Air-pouch experiments were performed as described previously⁷, using mice aged 8–12 weeks. In brief, 5 ml of 0.2- μ m-filtered air was injected subcutaneously into the dorsal region. After 3 days, the same pouches were injected with 3 ml of 0.2- μ m-filtered air to maintain the pouch. Four days later the pouches were injected with 1 ml of 0.2- μ m-filtered supernatants from control or ultraviolet-treated Jurkat cells or with lipopolysaccharide (1 μ g) or CCL2 (250 ng). After 24 h, cells from the air-pouch were collected by lavage twice with 2 ml of HBSS/1% FBS. Collected cells were resuspended in equal volumes and cell counting was performed with a haemocytometer and/or by flow cytometry. For analyses of specific populations in the air-pouch, cells were treated with anti-CD16/32 for 15 min on ice to block Fc receptor binding, followed by addition of the indicated fluorescently labelled antibodies for 30 min on ice. After being washed, the cells were analysed with a FACS Canto instrument (Becton Dickinson).

For thymic clearance studies, 4–5-week-old mice were injected intraperitoneally with 250 μ g dexamethasone in 300 μ l PBS. Apyrase-treated mice were injected with 5 U of recombinant apyrase at the midpoint of the dexamethasone treatment. Suramin-treated mice were injected intraperitoneally with 6 mg suramin in 300 μ l PBS 1 h before dexamethasone injection. In all cases, control groups were treated with equivalent volumes of vehicle controls. After treatment, the thymus was collected and dissociated over a 70- μ m mesh filter in cold HBSS/2% FBS and diluted 100-fold into 1 \times final annexin V binding buffer. Cells were incubated with annexin V and propidium iodide for 10 min at room temperature. AccuCount beads (Spherotech) were then added and the samples were analysed by flow cytometry (FACS Canto) in duplicate.

Migration assays. Transwell migration assays were performed by applying 100 μ l of THP-1 cells at 2×10^6 ml⁻¹ to the upper chamber in the same culture medium as the chemoattractant in the lower chamber (500 μ l) of a 24-well plate with 5- μ m pore-size Transwells (Corning) for 1 h at 37 °C and 5% CO₂. The percentage of migrated cells was determined by FACS with 5.1- μ m AccuCount beads and calculated as a percentage of the input cells. BMDMs were prepared as described previously⁷. Day 7 BMDM cells (5×10^4) were added to the upper well of a 5- μ m pore-size transwell and incubated for 2 h at 37 °C and 5% CO₂. The number of migrated cells on the underside of the membrane was determined by Diff-Quick staining and counting five random fields under $\times 10$ magnification per duplicate insert. For migration of primary human monocytes,

peripheral blood mononuclear cells from healthy donors were left to adhere to plastic dishes overnight at 37 °C and 5% CO₂. The next day, non-adherent cells were removed, and adherent cells (monocyte-enriched fraction) were collected, resuspended in RPMI/1% BSA/HEPES and applied to the upper chamber of a 3- μ m pore-size transwell at 5×10^4 cells per well for 2 h. Cells that migrated to the lower chamber were collected, counted and stained for CD14. The data presented reflect the relative number of CD14⁺ monocytes that migrated to the lower chamber, on the basis of post-migration CD14 staining.

siRNA and stable transfections. THP-1 cells (10^7) were transfected with 1 μ g of siRNA using the BTX Square Pulse T820 electroporator (one pulse at 250 V for 25 ms) and used 72 h after transfection. The pA-Puro-hCD39 plasmid was generated by PCR cloning of the human CD39 cDNA (Open Biosystems) into the pA-Puro vector. For generation of the stable line, cells were electroporated with 10 μ g of linearized plasmid (pA-Puro-hCD39) and selected in puromycin for 1 week before clonal expansion and screening for cell surface expression of CD39.

Nucleotide measurement. ATP was measured by the luciferin/luciferase assay by means of an LB953 AutoLumat luminometer (Berthold), as described previously²⁹. UTP concentrations were quantified by the UDP-glucose pyrophosphorylase-based reaction, as described³⁰. In brief, 100- μ l samples were incubated in the presence of 0.5 U ml⁻¹ UDP glucose pyrophosphorylase, 0.5 U ml⁻¹ inorganic pyrophosphatase, 1.6 mM CaCl₂, 2 mM MgCl₂, 25 mM HEPES, pH 7.4 and about 100,000 c.p.m. 1 μ M [¹⁴C]glucose 1-phosphate. Incubations were for 1 h at 30 °C. Reactions were terminated by heating the samples at 95 °C for 2 min. Conversion of [¹⁴C]glucose 1-phosphate to [¹⁴C]UTP was determined by high-performance liquid chromatography. All nucleotide measurements were conducted on supernatants prepared from cells cultured in RPMI/1% BSA/10 mM HEPES. For measurement of ATP without disturbing the cells, Jurkat cells were induced to undergo apoptosis in a 24-well plate and a 0.4- μ m-pore Transwell filter with 200 μ l of medium was submerged in the well. The amount of ATP that diffused through the Transwell was determined by acquiring samples from within the transwell at the indicated times.

Caspase activation and protease release. Caspase activation and protease release assays were performed with the Caspase-Glo and CytoTox-Glo (Promega) reagents, in accordance with the manufacturer's instructions.

Real-time PCR. cDNA was synthesized from 50 ng of DNase-treated RNA (RNeasy; Qiagen) with the use of Superscript III (Invitrogen). Real-time PCR was performed on the ABI Prism 7000 instrument with TaqMan probes (Applied Biosystems). Values shown are normalized to 18S levels.

Immunohistochemistry. Detection of apoptotic cells by Apostain (Alexis Biochemical) was performed as described previously²⁸.

Primer sequences. The siRNA used in Fig. 3b was purchased from Dharmacon: control siRNA, SMARTpool Control 1; P2Y₂ siRNA, SMARTpool human P2RY2 (a mixture of four oligonucleotide duplexes: 5'-CGAGAACACUAAGGACA UUUU-3' (sense) and 5'-AAUGUCCUAGUGUUCUGGUU-3' (antisense); 5'-CGACAGAACUGACAUGCAGUU-3' (sense) and 5'-CUGCAUGUCAGUUCUGUCGUU-3' (antisense); 5'-GGAAUGCGUCCACCACAUAUU-3' (sense) and 5'-UAUGUGGUGGACGCAUCCUU-3' (antisense); 5'-UGCCGCGUCUGGUCUAUUUU-3' (sense) and 5'-UAAUAGACCAGCAGCGGAUU-3' (antisense)).

Primers used for PCR detection of P2Y₂ by agarose gel in Fig. 3b were as follows: 5'-CTAAGCCAGCTACGGGAC-3' (forward) and 5'-TCCTATCC TCTGCATGTC-3' (reverse); real-time PCR primers, TaqMan primers from Applied Biosystems (sequence not available from the company).

Statistical analyses. Data are presented as means \pm s.e.m. Statistical significance for individual data points was determined by Student's two-tailed *t*-test. A *P* value less than 0.05 was considered statistically significant.

ErbB2 resembles an autoinhibited invertebrate epidermal growth factor receptor

Diego Alvarado¹, Daryl E. Klein¹ & Mark A. Lemmon¹

The orphan receptor tyrosine kinase ErbB2 (also known as HER2 or Neu) transforms cells when overexpressed¹, and it is an important therapeutic target in human cancer^{2,3}. Structural studies^{4,5} have suggested that the oncogenic (and ligand-independent) signalling properties of ErbB2 result from the absence of a key intramolecular ‘tether’ in the extracellular region that autoinhibits other human ErbB receptors, including the epidermal growth factor (EGF) receptor⁶. Although ErbB2 is unique among the four human ErbB receptors^{6,7}, here we show that it is the closest structural relative of the single EGF receptor family member in *Drosophila melanogaster* (dEGFR). Genetic and biochemical data show that dEGFR is tightly regulated by growth factor ligands⁸, yet a crystal structure shows that it, too, lacks the intramolecular tether seen in human EGFR, ErbB3 and ErbB4. Instead, a distinct set of autoinhibitory interdomain interactions hold unliganded dEGFR in an inactive state. All of these interactions are maintained (and even extended) in ErbB2, arguing against the suggestion that ErbB2 lacks autoinhibition. We therefore suggest that normal and pathogenic ErbB2 signalling may be regulated by ligands in the same way as dEGFR. Our findings have important implications for ErbB2 regulation in human cancer, and for developing therapeutic approaches that target novel aspects of this orphan receptor.

Ligand-induced activation of EGFR involves a marked change in the extracellular region from a ‘tethered’ (inactive) to an ‘extended’ (active) configuration⁹ (Fig. 1a) in which an exposed ‘dimerization arm’ in domain II drives the formation of receptor dimers^{10,11}. In tethered EGFR, the dimerization arm is occluded by autoinhibitory intramolecular interactions between domains II and IV, which are also seen in unliganded ErbB3 and ErbB4—but are absent in ErbB2 (refs 6, 12). ErbB2 is structurally unique. Even without a bound ligand its extracellular region resembles the extended (EGF-bound) form of EGFR (Fig. 1b, c), with the dimerization arm exposed and apparently ‘poised’ to drive receptor–receptor interactions^{4,5}. No known soluble ligand directly regulates ErbB2, and it is the only family member that transforms cells when simply overexpressed (without ligand addition)¹. Thus, ErbB2 is regarded as an ‘auto-activated’ receptor that adopts a constitutively activated configuration that can form signalling-active heterodimers (or homodimers) without direct regulation by a growth factor. These properties are thought to explain how ErbB2 overexpression causes cancer⁷. Although ErbB2 is viewed as an oddity among human ErbB receptors, we show here that it is the closest structural relative of the single EGF receptor family member of *D. melanogaster* (dEGFR). Moreover, the structural features that initially suggested constitutive activation of ErbB2 seem important for dEGFR autoinhibition. Thus, ErbB2 shares more similarities with a possible ancestral EGF receptor than does human EGFR itself.

We determined the 2.7-Å X-ray crystal structure of the unliganded dEGFR extracellular region, encompassing domains I to IV

(Supplementary Table 1). *D. melanogaster* contains a single EGFR/ErbB-receptor, which is tightly regulated by four different ligands (Spitz, Gurken, Keren and Vein) in distinct developmental contexts⁸. Ligand binding is required for dEGFR activation in cultured cells^{13,14} and for strong dimerization of its isolated extracellular region *in vitro*¹³. Sequence analyses indicate that the overall domain arrangement in dEGFR is the same as in human ErbB receptors, except for an extra cysteine-rich domain (domain V, which is predicted to be similar to domains II and IV) at the carboxy terminus of the invertebrate EGFR extracellular region. Over domains I–IV (about 620 amino-acid residues long), dEGFR shares 39% sequence identity with human EGFR (hEGFR) and 35% with human ErbB2 (Supplementary Fig. 1). Because it is tightly regulated by ligands, we expected that an unliganded form of the dEGFR extracellular region (s-dEGFR) would adopt a tethered configuration similar to that seen in Fig. 1a for hEGFR. Instead, we found that s-dEGFR encompassing domains I–IV (s-dEGFR Δ V) is fully extended even in the absence of ligand (Fig. 1d), and closely resembles sErbB2 (Fig. 1c). The s-dEGFR Δ V dimerization arm is exposed, and the ligand-binding sites on domains I and III are in direct contact (Fig. 1d). A structural overlay of sErbB2 and s-dEGFR Δ V (Fig. 2a) shows them to be remarkably similar. Thus, the same configuration is seen for the inactive state of one ErbB receptor extracellular region (s-dEGFR Δ V without ligand) and another that is thought to be constitutively active (sErbB2).

Small-angle X-ray scattering (SAXS) studies excluded the possibility that crystal packing causes s-dEGFR Δ V to be extended. SAXS measurements of the maximum molecular dimension (D_{\max}), together with low-resolution molecular envelopes, allow clear distinction between extended and tethered configurations of ErbB receptor extracellular regions in solution¹⁵. D_{\max} for s-dEGFR Δ V in solution is 130 Å (Supplementary Table 2), equal to the value measured for sErbB2 (ref. 15) and 25–30 Å larger than values for the tethered human EGFR extracellular region (about 105 Å)¹⁵. Low-resolution molecular envelopes (Fig. 2b) also show that s-dEGFR Δ V is extended in solution. SAXS studies of complete s-dEGFR (with domain V) gave an average D_{\max} of 165 Å (Supplementary Table 2), indicating that domain V simply projects from the end of domain IV to extend the structure (Fig. 2b and Supplementary Fig. 2). Mutational studies provide further evidence for the absence of an autoinhibitory tether in dEGFR. The affinity of human EGFR for its ligands is increased when the domain II/IV tether is weakened with mutations or abolished by removing domain IV^{16,17} (Supplementary Fig. 3a). These mutations favour EGF binding by reducing the work required to relocate domains I and III for interaction with the same EGF molecule (and do not cause constitutive hEGFR activation^{16,18,19}). Equivalent substitutions or deletions in s-dEGFR do not enhance Spitz binding (Supplementary Fig. 3b), indicating that dEGFR has no equivalent domain II/IV tether. Thus, our crystallographic and solution studies show that the

¹Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, 809C Stellar-Chance Laboratories, 422 Curie Boulevard, Philadelphia, Pennsylvania 19104-6059, USA.

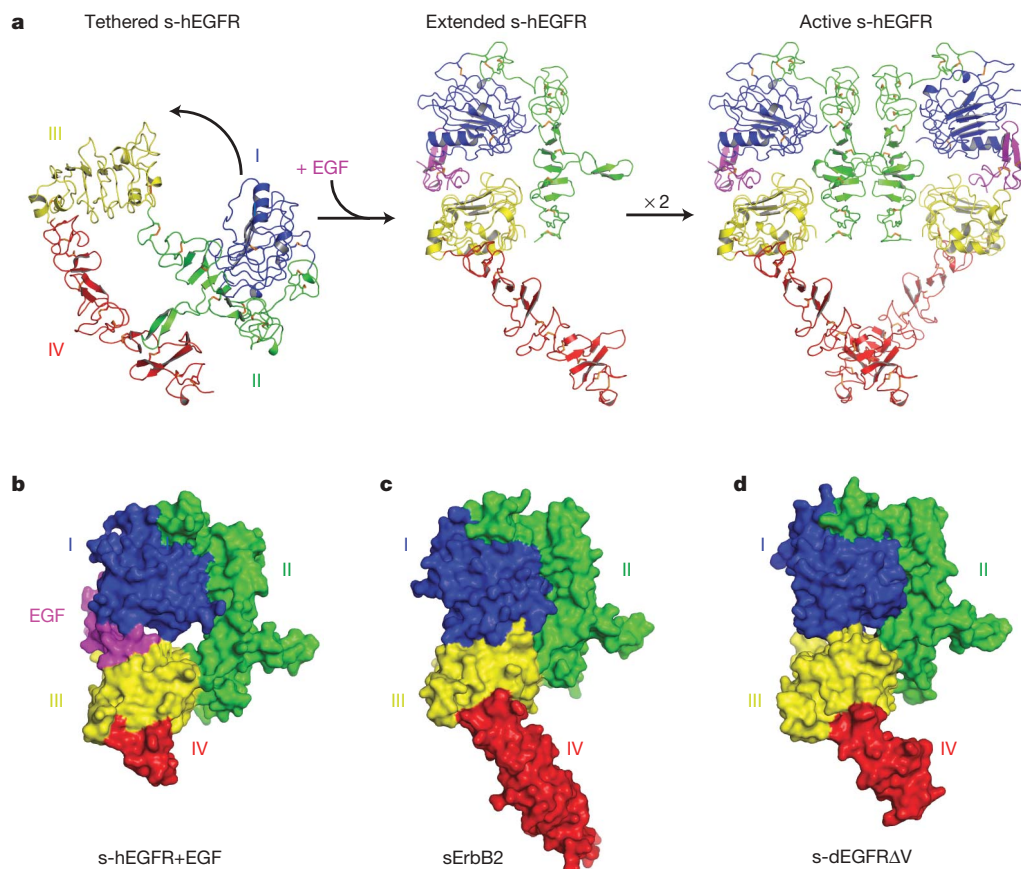


Figure 1 | Autoinhibition of ErbB receptors. **a**, The unliganded hEGFR extracellular region adopts a tethered structure (left), burying its dimerization arm (green) in autoinhibitory interactions between domains II and IV. Domains I, II, III and IV are blue, green, yellow and red, respectively. Binding of EGF (magenta) to domains I and III stabilizes extended s-hEGFR, exposing the dimerization arm (centre) to promote receptor dimerization (right)⁹. Most of domain IV was missing from extended s-hEGFR^{10,11}

unactivated *Drosophila* EGFR extracellular region adopts the same extended configuration as that seen for ErbB2.

Key elements of unliganded s-dEGFR overlay very well on the unactivated human EGFR extracellular region (s-hEGFR). As shown in Fig. 3a, the conformation of domain II in inactive s-dEGFRΔV (red) closely resembles that of domain II in inactive (tethered) s-hEGFR (grey) in an overlay using domain I as reference. This seems to be a characteristic 'inactive' domain II conformation, which is also shared by the unliganded ErbB3 and ErbB4 extracellular regions^{12,20}. By contrast, activated s-hEGFR¹¹ has a strikingly different domain II structure, with a roughly 12° bend between modules m4 and m5 (at the green arrow in Fig. 3b) that is known to be crucial for ligand-induced dimerization¹⁶. The domain II conformation in sErbB2 superimposes precisely on the inactive s-dEGFR and s-hEGFR structures (cyan structure in Fig. 3a), but not on the activated human EGFR structure. ErbB2 therefore has an 'inactive-like' domain II, indicating that published sErbB2 structures^{4,5} may represent an inactive (autoinhibited) configuration.

The failure of sErbB2 and unliganded s-dEGFRΔV to self-associate strongly, despite both having an exposed dimerization arm, also suggests an 'inactive', or dimerization-incompetent domain II conformation. The ErbB2 extracellular region does not homodimerize in solution^{21,22} or in crystals^{4,5}, and its heterodimerization with other sErbB proteins is barely detectable^{21,22}. Unliganded s-dEGFRΔV forms a crystallographic dimer mediated almost entirely by dimerization arm contacts (Supplementary Fig. 4). This self-association occurs only weakly in solution, with an approximate K_d of 40 μM as determined by analytical ultracentrifugation experiments (Supplementary Fig. 4).

structures and was added to the centre and right-hand panels using the domain IV structure of tethered s-hEGFR (left)¹⁷. **b**, Surface representation of a monomer from the EGF-bound s-hEGFR dimer (PDB accession 1ivo)¹¹. **c**, sErbB2 (PDB accession 1n8z, shown in surface representation) adopts an extended configuration similar to that of an activated s-hEGFR monomer⁴. **d**, Even in its inactive, unliganded state, s-dEGFRΔV is completely extended and closely resembles both sErbB2 and activated s-hEGFR.

Strong dimerization of s-dEGFRΔV or s-dEGFR requires Spitz binding (Supplementary Fig. 4a). Thus, the extracellular region of ErbB2—the human ErbB receptor believed to be unique in its ability to form ligand-independent homodimers and heterodimers^{7,23}—has less propensity for self-association than the equivalent region of the unliganded *Drosophila* EGF receptor. ErbB2 also shows no greater tendency than unliganded hEGFR to homodimerize in cells²⁴, and it is not constitutively active when expressed at physiologically relevant levels in insect cells²⁵. Together, these data point to ErbB2 being an inactive receptor—and one that may be more stringently autoinhibited than dEGFR.

Figure 3 suggests a mechanism for dEGFR regulation by growth factor binding that may also be relevant for ErbB2. Wedging an EGF-like molecule between the two ligand-binding domains I and III will push them apart as shown in Fig. 3c, necessitating a significant bend in domain II (which links domains I and III). Movement of disulphide-bonded module m5 with respect to m4 (at the green arrow in Fig. 3b) accounts for most of this bend, and effectively links ligand binding to reorientation of the dimerization arm. The result is a bent domain II conformation that can present a self-complementary dimerization interface (for homodimerization) or one that is optimized for heterodimerization. Direct interactions between domains I and III of s-dEGFR work against this process, and are therefore autoinhibitory. Interactions between domains I and III of s-dEGFRΔV involve regions that correspond exactly to the ligand-binding sites of hEGFR^{10,11}, and they therefore directly occlude the ligand-binding sites (Fig. 1a)—burying 452 Å² of surface. Details of these interactions are shown in Supplementary Fig. 5a. The same elements in sErbB2 also contribute

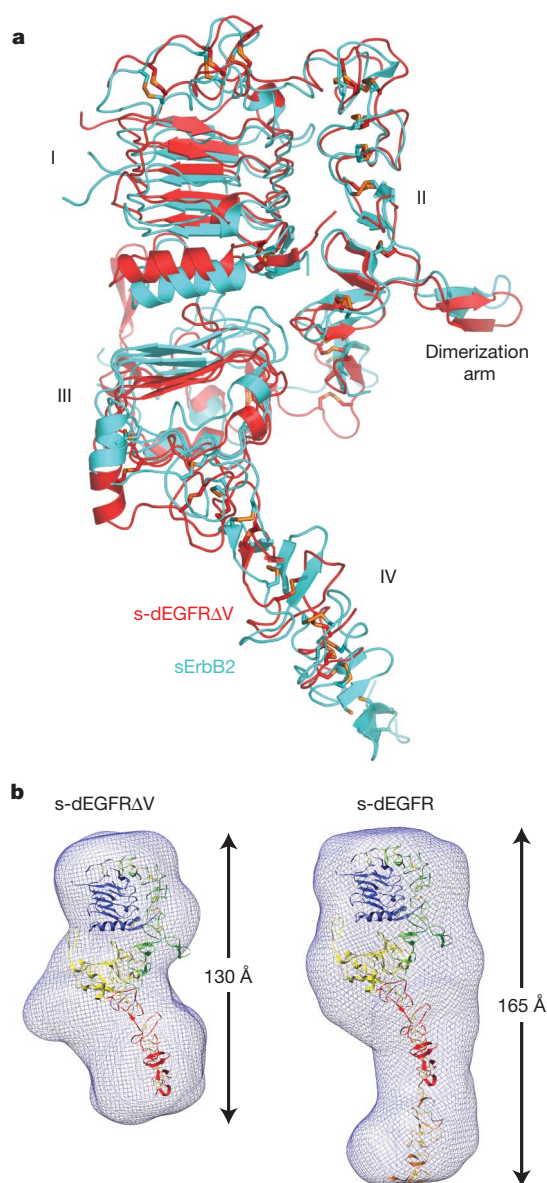


Figure 2 | The unactivated dEGFR extracellular region closely resembles sErbB2. **a**, Global superimposition of inactive s-dEGFR Δ V (red) and sErbB2 (cyan)⁴ illustrates their conformational similarity. Direct interactions between domains I and III (more extensive in sErbB2 than in s-dEGFR) help to stabilize the extended configuration in both receptors (Supplementary Fig. 5) and block ligand-binding sites. **b**, Low-resolution molecular envelopes from SAXS studies of s-dEGFR Δ V (left) and s-dEGFR (right), with maximum molecular dimensions (D_{\max}) marked (see Supplementary Table 2). The s-dEGFR Δ V envelope readily accommodates the crystallographic model. In intact s-dEGFR, domain V (orange) seems simply to add to the maximum dimension. Domain V and the C terminus of domain IV (poorly defined in our crystal structure) were modelled with s-hEGFR domain IV as template. In the right-hand panel, the three most C-terminal disulphide-linked modules of domain V have been removed. The fact that these are not accommodated by the SAXS envelope suggests flexibility at the C terminus.

to direct interactions between domains I and III^{4,5} but are augmented by additional contacts to bring domains I and III even closer together (by about 8 Å) than in s-dEGFR (Fig. 2a), burying a total surface of about 1,250 Å² (Supplementary Fig. 5b). The direct interactions between domains I and III seen in dEGFR (and ErbB2) are autoinhibitory because they force the two parts of the ligand-binding site so close to one another that ligand cannot be accommodated. By

contrast, the domain II/IV tether in hEGFR (Fig. 1a) pulls the two halves of the ligand-binding site (on domains I and III) too far apart for them both to contact the same ligand molecule simultaneously. The autoinhibitory consequence for ligand binding is similar in both cases, with work being required to separate domains I and III in dEGFR but (conversely) to draw them together in hEGFR by breaking the domain II/IV tether. Thus, these are variations on the same autoinhibitory theme.

The close apposition of domains I and III in dEGFR also promotes an important set of interactions between domains I and II (Fig. 3c) that stabilize the inactive domain II conformation. Side chains from the 'back' of s-dEGFR domain II in modules m5 and m6 (Y259 and H270, respectively) pack against a hydrophobic patch on domain I comprising the side chains of I2, I4 and Y32 (Fig. 3c and Supplementary Fig. 6a) and form hydrogen bonds with D34 in domain I. These interactions restrain the orientation of modules m5 and m6 with respect to m4 and maintain the dimerization arm in the 'inactive' position shown in Fig. 3a, c. Very similar sets of interactions between domains I and II occur in sErbB2 (Supplementary Fig. 6b), in inactive hEGFR (Supplementary Fig. 6c), and also in unliganded ErbB3 and ErbB4 (refs 12, 20). All of these interactions are broken in the active configuration (Fig. 3d and Supplementary Fig. 6d), so that domain II modules m5 and m6 no longer make direct contact with domain I, and the dimerization arm becomes reoriented. Disrupting these interactions between domains I and II in s-dEGFR, by mutation of Y259 and H270 to alanine and serine, respectively, enhances Spitz binding to about the same extent as domain II/IV tether mutations enhance EGF binding to s-hEGFR¹⁷ (Supplementary Fig. 7 and Supplementary Table 3). From the perspective of ligand binding, domain I/II contacts in dEGFR therefore constitute an (autoinhibitory) energetic barrier that is similar in strength to the domain II/IV tether in human EGFR, ErbB3 and ErbB4. Disrupting domain I/II contacts in intact dEGFR or ErbB2 did not elevate the constitutive activity of these receptors (data not shown), but neither does disruption of domain II/IV tether contacts in hEGFR^{16,18,19}. Breaking autoinhibitory interactions in the extracellular region, although necessary for activation, is clearly not sufficient. Indeed, even if domain IV is deleted entirely from s-hEGFR (so that the tether cannot form), dimerization still requires EGF addition¹⁰. Thus, the unliganded *Drosophila* and human EGF receptors rely on different sets of autoinhibitory intramolecular interactions to oppose ligand binding and dimerization.

The fact that ErbB2 maintains—and even extends—all of the autoinhibitory interactions seen in *Drosophila* EGFR militates against the prevailing notion that ErbB2 is 'poised' to dimerize through its exposed dimerization arm^{4,5}. Furthermore, the failure of sErbB2 to form homodimers or heterodimers *in vitro*^{21,22} suggests that it is even more stringently autoinhibited than s-dEGFR (which does homodimerize weakly)—which is consistent with its larger domain I/III interface (Supplementary Fig. 5). Nonetheless, crosslinking and co-immunoprecipitation studies show that intact ErbB2 can form homodimers and heterodimers in mammalian cells^{7,23,26}. One possible explanation is that ErbB2 relies uniquely on interactions outside its extracellular region to drive dimerization. A second possibility is that unknown cellular ligands promote ErbB2 activation when it is over-expressed in mammalian cells (but not in insect cells²⁵). The first of these possibilities is countered by reports that deletion of the cytoplasmic region does not abolish ErbB2 homodimerization or heterodimerization²⁶—although a key role for the transmembrane domain cannot be excluded. The second possibility seems unlikely, given the failure of substantial efforts in the 1980s and 1990s to identify soluble ligands that directly activate ErbB2 (ref. 7).

Although no genuine soluble ligand for ErbB2 is known, at least one membrane-bound regulator that contains EGF-like domains has been identified²⁷. A subunit of Muc4 (ASGP2) was reported to interact with ErbB2 and promote its tyrosine phosphorylation. An EGF-like domain in membrane-associated Muc4 might bind between domains I and III of ErbB2 and induce conformational changes of

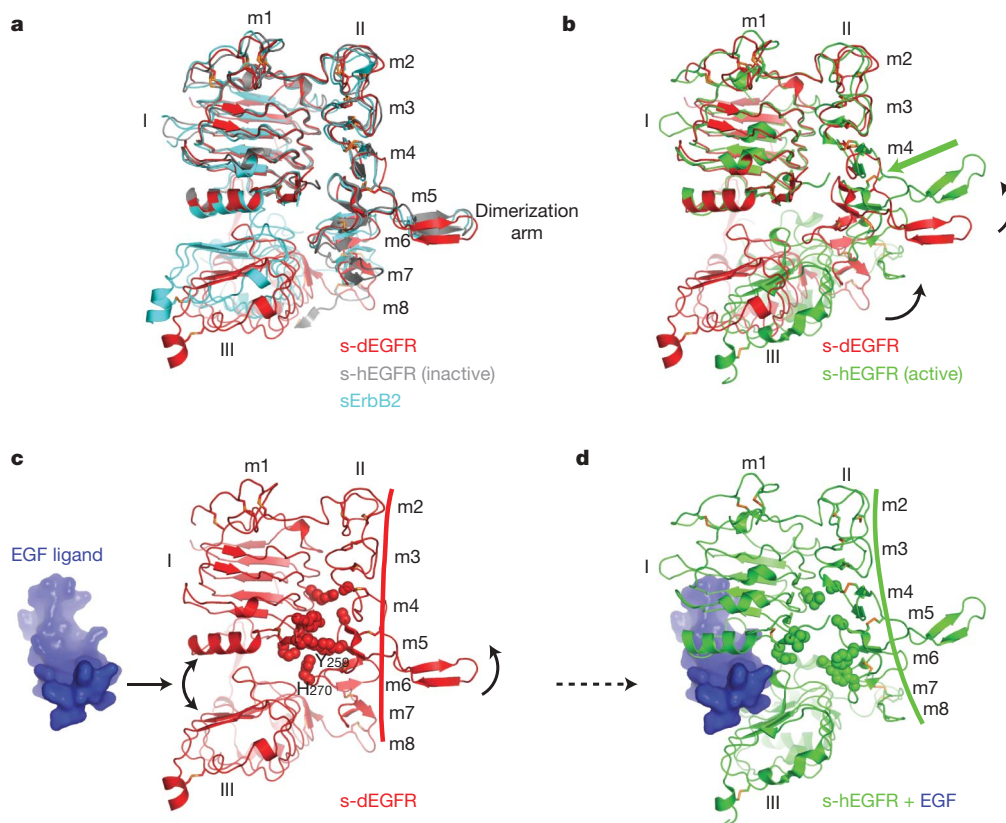


Figure 3 | Ligand binding breaks autoinhibitory interactions between domains I and II common to s-dEGFR, s-hEGFR and sErbB2.

a, Superposition of inactive s-hEGFR (grey) on s-dEGFR (red) and sErbB2 (cyan), with domain I as reference. The eight disulphide-bonded modules (m1–m8) that define domain II are labelled, as is the dimerization arm—located almost identically in all three structures. Domain III of inactive s-hEGFR has been removed for clarity. **b**, A similar overlay of active s-hEGFR (green) and inactive s-dEGFR (red) highlights the reorientation of the dimerization arm on ligand binding. The structures overlay very well

the sort depicted in Fig. 3c, d to promote the ability of ErbB2 to form homodimers and/or heterodimers. Similarly, it has been shown genetically in *Drosophila* that Spitz must be palmitoylated (which drives its membrane association²⁸) to regulate dEGFR *in vivo*²⁸. Gurken and Keren have a similar palmitoylation site, whereas Vein—considered to be a ‘weak’ dEGFR ligand⁸—does not. Thus, membrane association seems to be a key feature of ligands (Muc4 and Spitz) that activate the two ErbB receptors known to adopt an extended configuration in the absence of ligand (ErbB2 and dEGFR). Membrane association may be required to increase the local ligand concentration at the cell surface, so as to promote the ligand’s ability to ‘wedge apart’ domains I and III of dEGFR or ErbB2 (breaking autoinhibitory interactions between domains I and III). By contrast, the tethered configuration of hEGFR, ErbB3 and ErbB4 (Fig. 1a) keeps the ligand-binding sites on domains I and III fully exposed and freely accessible to soluble growth factors. We speculate that evolution of the domain II/IV tether as a distinct mode of autoinhibition might have occurred alongside the emergence of the ability of ErbB receptors (other than ErbB2) to respond to soluble (rather than membrane-bound) growth factor ligands.

Given the importance of ErbB2 in human cancer, and its validated utility as a target of cancer therapeutics³, the view of ErbB2 regulation presented here has several implications. The development of agents that stabilize autoinhibitory interactions might represent a new therapeutic avenue for inhibiting ErbB2 signalling in cancer. Equally, the fact that ErbB2 shows such striking resemblance to a tightly

in modules m1–m4 of domain II, but deviate significantly at the m4–m5 linkage (green arrow) because of a ligand-induced bend. **c**, **d**, Model for activation of dEGFR (and ErbB2) by wedging an EGF-like ligand (blue) between domains I and III. Forcing domains I and III apart (**c**) disrupts all direct interactions between domains I and III, as well as a set of domain I/II contacts that normally maintain domain II in an inactive conformation (residues shown in space-filling representation: see Supplementary Fig. 6). In EGF-bound s-hEGFR (**d**), the side chains shown in green space-filling representation no longer interact, and domain II is bent.

ligand-regulated invertebrate EGF receptor suggests that ErbB2 also has activating ligands. Identifying these probably membrane-associated ligands, and understanding their role in activating ErbB2 in different human cancers, should provide new directions for therapeutic targeting of ErbB receptor signalling.

METHODS SUMMARY

Histidine-tagged s-dEGFR and s-dEGFRΔV were produced by secretion from baculovirus-infected *Spodoptera frugiperda* Sf9 cells or transfected *Drosophila* S2 cells. The C-terminal amino acid of s-dEGFRΔV was T589 in the numbering convention used in Supplementary Fig. 1 (see Methods). Secreted protein was collected by metal-affinity chromatography and further purified by ion-exchange and size-exclusion chromatography as described¹³. Surface plasmon resonance (SPR), SAXS and sedimentation equilibrium analytical ultracentrifugation studies were performed essentially as described^{13,15,16}.

Purified s-dEGFRΔV was crystallized using the vapour diffusion method in 10% PEG 4000, 5% Jeffamine M-600, pH 7.0, 12.5% ethylene glycol, 100 mM HEPES, pH 7.4, 50 mM KCl. Plate-shaped crystals of approximate dimensions 200 μm × 200 μm × 75 μm grew in 1–5 days and were frozen directly from the mother liquor. Data were collected using beamline 23ID-D at the Advanced Photon Source (Argonne, Illinois) as described in Supplementary Table 1. The structure of s-dEGFRΔV was solved using molecular replacement (MR) methods. Search models based on the coordinates of domains I and III from ErbB2 (PDB accession 2a91)⁵ were generated by replacing non-conserved amino acids with alanines. Although MR solutions could not be found for domains II or IV, initial maps calculated using phases from models containing only domains I and III showed strong density for domain II. Model building with COOT²⁹ was alternated with successive rounds of restrained refinement using REFMAC³⁰. In

later stages of refinement, composite omit maps were generated, which allowed much of domain IV to be built and oligosaccharides to be placed.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 April; accepted 16 July 2009.

Published online 30 August 2009.

- Di Fiore, P. P. *et al.* erbB-2 is a potent oncogene when overexpressed in NIH/3T3 cells. *Science* **237**, 178–182 (1987).
- Hynes, N. E. & Lane, H. A. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nature Rev. Cancer* **5**, 341–354 (2005).
- Moasser, M. M. Targeting the function of the HER2 oncogene in human cancer therapeutics. *Oncogene* **26**, 6577–6592 (2007).
- Cho, H. S. *et al.* Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature* **421**, 756–760 (2003).
- Garrett, T. P. *et al.* The crystal structure of a truncated ErbB2 ectodomain reveals an active conformation, poised to interact with other ErbB receptors. *Mol. Cell* **11**, 495–505 (2003).
- Burgess, A. W. *et al.* An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Mol. Cell* **12**, 541–552 (2003).
- Citri, A., Skaria, K. B. & Yarden, Y. The deaf and the dumb: the biology of ErbB-2 and ErbB-3. *Exp. Cell Res.* **284**, 54–65 (2003).
- Shilo, B. Z. Regulating the dynamics of EGF receptor signaling in space and time. *Development* **132**, 4017–4027 (2005).
- Lemmon, M. A. Ligand-induced ErbB receptor dimerization. *Exp. Cell Res.* **315**, 638–648 (2009).
- Garrett, T. P. *et al.* Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor α . *Cell* **110**, 763–773 (2002).
- Ogiso, H. *et al.* Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* **110**, 775–787 (2002).
- Bouyain, S., Longo, P. A., Li, S., Ferguson, K. M. & Leahy, D. J. The extracellular region of ErbB4 adopts a tethered conformation in the absence of ligand. *Proc. Natl Acad. Sci. USA* **102**, 15024–15029 (2005).
- Klein, D. E., Nappi, V. M., Reeves, G. T., Shvartsman, S. Y. & Lemmon, M. A. Argos inhibits epidermal growth factor receptor signalling by ligand sequestration. *Nature* **430**, 1040–1044 (2004).
- Schweitzer, R., Shaharabany, M., Seger, R. & Shilo, B. Z. Secreted Spitz triggers the DER signaling pathway and is a limiting component in embryonic ventral ectoderm determination. *Genes Dev.* **9**, 1518–1529 (1995).
- Dawson, J. P., Bu, Z. & Lemmon, M. A. Ligand-induced structural transitions in ErbB receptor extracellular domains. *Structure* **15**, 942–954 (2007).
- Dawson, J. P. *et al.* Epidermal growth factor receptor dimerization and activation require ligand-induced conformational changes in the dimer interface. *Mol. Cell Biol.* **25**, 7734–7742 (2005).
- Ferguson, K. M. *et al.* EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. *Mol. Cell* **11**, 507–517 (2003).
- Mattoon, D., Klein, P., Lemmon, M. A., Lax, I. & Schlessinger, J. The tethered configuration of the EGF receptor extracellular domain exerts only a limited control of receptor function. *Proc. Natl Acad. Sci. USA* **101**, 923–928 (2004).
- Walker, F. *et al.* CR1/CR2 interactions modulate the functions of the cell surface epidermal growth factor receptor. *J. Biol. Chem.* **279**, 22387–22398 (2004).
- Cho, H. S. & Leahy, D. J. Structure of the extracellular region of HER3 reveals an interdomain tether. *Science* **297**, 1330–1333 (2002).
- Ferguson, K. M., Darling, P. J., Mohan, M. J., Macatee, T. L. & Lemmon, M. A. Extracellular domains drive homo- but not hetero-dimerization of erbB receptors. *EMBO J.* **19**, 4632–4643 (2000).
- Horan, T. *et al.* Binding of Neu differentiation factor with the extracellular domain of Her2 and Her3. *J. Biol. Chem.* **270**, 24604–24608 (1995).
- Penuel, E., Akita, R. W. & Sliwkowski, M. X. Identification of a region within the ErbB2/HER2 intracellular domain that is necessary for ligand-independent association. *J. Biol. Chem.* **277**, 28468–28473 (2002).
- Wehrman, T. S. *et al.* A system for quantifying dynamic protein interactions defines a role for Herceptin in modulating ErbB2 interactions. *Proc. Natl Acad. Sci. USA* **103**, 19063–19068 (2006).
- Berger, M. B., Mendrola, J. M. & Lemmon, M. A. ErbB3/HER3 does not homodimerize upon neuregulin binding at the cell surface. *FEBS Lett.* **569**, 332–336 (2004).
- Brennan, P. J., Kumagai, T., Berezov, A., Murali, R. & Greene, M. I. HER2/neu: mechanisms of dimerization/oligomerization. *Oncogene* **19**, 6093–6101 (2000).
- Carraway, K. L., Ramsauer, V. P., Haq, B. & Carothers Carraway, C. A. Cell signaling through membrane mucins. *BioEssays* **25**, 66–71 (2003).
- Miura, G. I. *et al.* Palmitoylation of the EGFR ligand Spitz by Rasp increases Spitz activity by restricting its diffusion. *Dev. Cell* **10**, 167–176 (2006).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- CCP4 (Collaborative Computational Project Number 4). The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Lemmon and Ferguson laboratories, K. Ferguson, G. Van Duyne, C. Abrams and J. Shorter for advice and comments on the manuscript; Z. Bu for assistance with collecting SAXS data; and R. Gillilan of MacCHESS for help with SAXS data collection and processing at CHESS beamline G1. Crystallographic results were obtained in research conducted at the GM/CA Collaborative Access Team at the Advanced Photon Source (APS), supported with funds from the National Cancer Institute and National Institute of General Medical Science. This study was supported by grants from the NIH (to M.A.L.). D.E.K. was supported by a Predoctoral fellowship from the US Army Breast Cancer Research Program. D.A. was supported by an NIH Postdoctoral Training Grant and a Postdoctoral Fellowship from the Damon Runyon Cancer Research Foundation.

Author Contributions D.A., D.E.K. and M.A.L. conceived and designed the project. D.E.K. established initial expression and purification procedures for s-dEGFR variants and relevant ligands. D.A. was responsible for designing s-dEGFR constructs used in this study and executed all biophysical studies, crystallization and data collection. D.A. also solved and refined the s-dEGFR Δ V structure. D.A. and M.A.L. interpreted data and wrote the manuscript.

Author Information The structure for s-dEGFR Δ V has been deposited in the Protein Data Bank under accession number 3I2T. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.A.L. (mlemmon@mail.med.upenn.edu).

METHODS

Protein expression and purification. Coding regions for wild-type and mutated forms of s-DEGFR were subcloned into pFastbac-1 and pMT/V5-His A (Invitrogen) for expression in *Spodoptera frugiperda* (Sf9) and *Drosophila melanogaster* Schneider-2 (S2) cells respectively. A C-terminal hexahistidine tag was incorporated into all constructs by PCR. s-DEGFRΔV ended at T589 using the numbering scheme employed in Supplementary Fig. 1 (see the comments in Crystallography section below on dEGFR numbering), and s-DEGFRΔIV-V was truncated at N493. Two sets of mutations were made to disrupt the domain I/II autoinhibitory interface for Supplementary Fig. 7 and Supplementary Table 3. In one, Y259 and H270 in domain II were mutated to alanine and serine, respectively. In the second, sites in domains I and II were mutated to give the tetramutant I2A/Y32A/Y259A/H270S. The effects of these mutations were assessed both in the background of wild-type s-DEGFR and a Y242S/Y247S mutant in which dimer contacts had been disrupted. The s-DEGFR^{dim-arm} construct referred to in Supplementary Fig. 4 and Supplementary Table 2 contains a series of mutations in the domain II dimerization arm analogous to those previously shown to abolish ligand-induced dimerization of human sEGFR¹⁰: Y242E, N243A, T245D, Y247E, V248A and L249D. The s-DEGFR^{ther} mutant referred to in Supplementary Fig. 3 contains three mutations in domain IV analogous to those that break all intramolecular hydrogen-bonding interactions between domains II and IV observed in the unliganded s-hEGFR structure¹⁷: D547A, H550A and K559A. All mutations were generated with the QuikChange mutagenesis kit (Stratagene) and fully sequenced.

Stable S2 cell pools and recombinant baculoviruses were generated as described^{13,17,31}, and each protein of interest was secreted into the culture medium. For S2 cell-expressed proteins, S2 cells were grown in EX-CELL 420 serum-free medium (Sigma-Aldrich) to a density of about $(5\text{--}10) \times 10^6$ cells ml⁻¹, and protein expression was induced with 500 μM CuSO₄ for 3–4 days. For Sf9 cell-expressed proteins, Sf9 cells were grown in Sf900II medium (Invitrogen-Gibco) to a density of $(2\text{--}3) \times 10^6$ cells ml⁻¹ and were infected with recombinant baculovirus for 3–4 days. In each case, 2–4 l of conditioned medium were flowed over a 3–4-ml bed volume of Ni²⁺-nitrilotriacetate agarose (Qiagen). After the column had been washed with 25 mM MES, pH 6.0, 150 mM NaCl (buffer A), bound proteins were eluted with increasing concentrations of imidazole in buffer A. Protein-containing fractions were applied to a Uno-S (Bio-Rad) cation-exchange column equilibrated in buffer A, and were eluted with a salt gradient from 150 mM to 1 M NaCl in buffer A. s-DEGFR proteins were eluted between 200 and 500 mM NaCl and were concentrated with a Centricon-50 concentrator (Millipore) before further purification by size-exclusion chromatography with a Superose-6 column (GE Healthcare) equilibrated in buffer B (25 mM HEPES, pH 8.0, 150 mM NaCl). Secreted Spitz and Spitz^{C29S} were purified from S2 cells exactly as described previously^{13,28,31}.

Surface plasmon resonance (SPR). Secreted Spitz was immobilized on CM5 sensorchips by using amine coupling, exactly as described previously¹³. Increasing concentrations of s-DEGFR proteins (12.5–6,400 nM) were then flowed over the sensorchip in buffer B at 25 °C. The sensorchip surface was regenerated after each injection with the use of 10 mM sodium acetate, pH 4.5, 1 M NaCl, as described¹³. The maximum SPR response at steady state for each s-DEGFR concentration was plotted against s-DEGFR concentration, and the resulting curves could be fitted straightforwardly to simple binding isotherms with the program Prism (GraphPad), from which apparent K_d values were obtained. Standard error of the mean values were generated from at least three independent measurements, using at least two independent preparations of each protein.

SAXS. SAXS data were collected at 25 °C with a rotating anode source at Fox Chase Cancer Center, as described¹⁵, or at CHESS beamline G1, using protein samples at concentrations between 1 and 6 mg ml⁻¹ in buffer B. Data handling and reduction were performed as described previously¹⁵ or with the program Datasqueeze (Datasqueeze Software). Potential problems with radiation-induced denaturation were monitored by inspection of Kratky plots with increasing exposure time, graphing IQ^2 as a function of Q , where I is the scattered intensity and $Q = 4\pi\sin(\theta/2)/\lambda$ (where Q is the magnitude of the scattering vector, θ the scattering angle and λ the X-ray wavelength). The program GNOM³² was used to obtain $P(r)$ curves, the maximum dimension of the molecule (D_{\max}) and its radius of gyration (R_g). Quoted R_g values (Supplementary Table 2) represent means (and standard deviations) from at least three independent determinations. D_{\max} values were determined empirically by recomputing $P(r)$ curves in GNOM with a series of different r_{\max} values (in steps of 5 Å), and selecting as D_{\max} the r_{\max} value at which $P(r)$ most closely approached zero while giving a plausible $P(r)$ curve. Errors in D_{\max} values are quoted as ± 5 Å on the basis of the empirical approach used for their determination. Low-resolution molecular envelopes were generated *ab initio* with the program DAMMIN as described previously^{15,33}, using SAXS data collected on the home source with s-DEGFR concentrations lower than 1 mg ml⁻¹. In brief, ten iterations of DAMMIN were averaged and filtered as described³⁴, using

the DAMAVER suite of programs. Crystal structures of models were docked into the resulting 'most probable' envelopes with SITUS³⁵, and the outputs were displayed and manually refined with the UCSF Chimera package (<http://www.cgl.ucsf.edu/chimera>)³⁶.

Sedimentation equilibrium ultracentrifugation. Experiments were performed exactly as described²¹, with the following modifications. Receptor extracellular regions at 2, 4 and 8 μM, both with and without a 1.2-fold excess of Spitz, were centrifuged in buffer B at 6,000, 9,000 and 12,000 r.p.m. in an Optima XL-A analytical ultracentrifuge (Beckman) at 20 °C, using absorbance at 280 nm to detect protein distribution. The program Winmatch (<http://www.biotech.uconn.edu/auf/>) was used to ensure that samples had reached equilibrium. Data were analysed with Sedfit and Sedphat (<http://www.analyticalultracentrifugation.com>) and were fitted to a monomer–dimer equilibrium model as described¹⁶, considering s-DEGFR or the s-DEGFR–Spitz complex as the dimerizing species. Fits used to determine the quoted K_d values gave good residuals, with no systematic deviations. In Supplementary Fig. 4a, sedimentation data are plotted as $\ln A_{280}$ against $(r^2 - r_0^2/2)$, where r is the radial position in the sample and r_0 is the radial position of the meniscus. For a single species this representation gives a straight line with a slope proportional to its molecular mass. Standard deviations quoted represent data from at least three independent experiments.

Crystallography. *Generation of s-DEGFRΔV protein suitable for crystallization.* The *torpedo* locus in *Drosophila melanogaster* encodes two splice variants named dEGFR1 and dEGFR2 that differ only at their N termini^{37,38}. Mature dEGFR1 and dEGFR2 have N-terminal extensions of 21 and 71 amino acids, respectively, which show no significant sequence similarity, are devoid of significant predicted secondary structure and are proteolytically labile as determined by N-terminal sequencing of the corresponding s-DEGFR species. We found that s-DEGFR1 and s-DEGFR2 bind Spitz with the same affinity (data not shown) and that removal of the N-terminal extensions has no influence on Spitz binding (data not shown). Beyond amino acid 22 of mature dEGFR1 (C53 of predicted pro-dEGFR1) and amino acid 72 of mature dEGFR2 (C102 of pro-dEGFR2), the two splice forms are identical. Therefore, to generate a protein amenable to crystallization (s-DEGFR), we deleted amino acids 1–21 and 1–71 of mature dEGFR1 and dEGFR2, respectively (equivalent to amino acids 1–52 and 1–101 of the respective pro forms), so that the N-terminal amino acid of mature s-DEGFR corresponds to K20 of mature dEGFR1 or K70 of mature dEGFR2 (the second residue—V21 in mature dEGFR1 and I71 in mature dEGFR2—is I2 in mature s-DEGFR; Supplementary Fig. 1). Immediately before K1 of s-DEGFR, we added a BiP signal sequence (substituted for the native one) to drive the secretion of s-DEGFR into the culture medium, followed by a hexahistidine tag (in addition to the His₆ tag at the C terminus), so that the presumed mature s-DEGFR protein secreted from S2 or Sf9 cells starts with six histidines, which we number –5 to 0. Domain V was also deleted from s-DEGFR at T589 (using the numbering in Supplementary Fig. 1), yielding s-DEGFRΔV, which also has a C-terminal hexahistidine tag. Whereas crystals grown with s-DEGFR2 protein diffracted poorly, s-DEGFRΔV crystals diffracted well to 2.7 Å resolution.

Crystallization and data collection. Purified s-DEGFRΔV (see above) at 100 μM was crystallized with the vapour diffusion method by mixing equal volumes of protein with a solution containing 10% PEG 4000, 5% Jeffamine M-600, pH 7.0, 12.5% ethylene glycol, 100 mM HEPES, pH 7.4, 50 mM KCl, and equilibrating the mixture over a reservoir of this solution at 21 °C. Plate-shaped crystals of approximate dimensions 200 μm × 200 μm × 75 μm grew in 1–5 days and were frozen directly from the mother liquor. Data were collected using beamline 23ID-D at the Advanced Photon Source (Argonne, Illinois) and were processed with HKL-2000 (ref. 39). Crystals were of space group C22₁, with unit cell dimensions $a = 74.4$ Å, $b = 174.8$ Å, $c = 161.6$ Å and $\alpha = \beta = \gamma = 90^\circ$. There is one s-DEGFRΔV molecule in the asymmetric unit, with a Matthews coefficient of $3.2 \text{ Å}^3 \text{ Da}^{-1}$, giving a solvent content of 62.2%.

Molecular replacement and refinement. The structure of s-DEGFRΔV was solved with MR methods. Search models based on the coordinates of domains I and III from ErbB2 (PDB accession 2a91)⁵ were generated by replacing non-conserved amino acids with alanines. Domains I and III were found in simultaneous but independent searches by using PHASER (CCP4)³⁰. Although we were unable to find MR solutions for domains II or IV with a variety of search models, initial maps based on domain I/III models showed strong density for domain II. Model building with COOT²⁹ was alternated with successive rounds of restrained refinement with REFMAC³⁰ and solvent flattening with DM³⁰. In later stages of refinement, composite omit maps were generated in CNS⁴⁰, which allowed much of domain IV to be built and oligosaccharides to be placed. The final stages of refinement employed TLS refinement⁴¹ with anisotropic motion tensors refined for each of the four domains, using REFMAC³⁰.

Calculations and figure preparation. Calculations of buried surface were performed with AREAIMOL in the CCP4 suite of programs³⁰. Calculations of surface complementarity, S_c (ref. 42), used the program SC in CCP4 (ref. 30).

Structure validation was performed with SFCHECK and PROCHECK in CCP4 (ref. 30). Figures were generated with MacPymol⁴³ (<http://www.pymol.org>).

31. Klein, D. E., Stayrook, S. E., Shi, F., Narayan, K. & Lemmon, M. A. Structural basis for EGFR ligand sequestration by Argos. *Nature* **453**, 1271–1275 (2008).
32. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).
33. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886 (1999).
34. Volkov, V. V. & Svergun, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864 (2003).
35. Wriggers, W. & Birmanns, S. Using Situs for flexible and rigid-body fitting of multiresolution single-molecule data. *J. Struct. Biol.* **133**, 193–202 (2001).
36. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
37. Clifford, R. & Schupbach, T. Molecular analysis of the *Drosophila* EGF receptor homolog reveals that several genetically defined classes of alleles cluster in subdomains of the receptor protein. *Genetics* **137**, 531–550 (1994).
38. Lesokhin, A. M., Yu, S. Y., Katz, J. & Baker, N. E. Several levels of EGF receptor signaling during photoreceptor specification in wild-type, *Ellipse*, and null mutant *Drosophila*. *Dev. Biol.* **205**, 129–144 (1999).
39. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
40. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
41. Winn, M. D., Isupov, M. N. & Murshudov, G. N. Use of TLS anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D* **57**, 122–133 (2001).
42. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
43. DeLano, W. L. *The PyMOL Molecular Graphics System* (DeLano Scientific, 2002).

LETTERS

Structure of the BK potassium channel in a lipid membrane from electron cryomicroscopy

Liguo Wang¹ & Fred J. Sigworth¹

A long-sought goal in structural biology has been the imaging of membrane proteins in their membrane environments. This goal has been achieved with electron crystallography¹ in those special cases where a protein forms highly ordered arrays in lipid bilayers. It has also been achieved by NMR methods¹ in proteins up to 50 kilodaltons (kDa) in size, although milligram quantities of protein and isotopic labelling are required. For structural analysis of large soluble proteins in microgram quantities, an increasingly powerful method that does not require crystallization is single-particle reconstruction from electron microscopy of cryogenically cooled samples (electron cryomicroscopy (cryo-EM))². Here we report the first single-particle cryo-EM study of a membrane protein, the human large-conductance calcium- and voltage-activated potassium channel³ (BK), in a lipid environment. The new method is called random spherically constrained (RSC) single-particle reconstruction. BK channels, members of the six-transmembrane-segment (6TM) ion channel family, were reconstituted at low density into lipid vesicles (liposomes), and their function was verified by a potassium flux assay. Vesicles were also frozen in vitreous ice and imaged in an electron microscope. From images of 8,400 individual protein particles, a three-dimensional (3D) reconstruction of the BK channel and its membrane environment was obtained at a resolution of 1.7–2.0 nm. Not requiring the formation of crystals, the RSC approach promises to be useful in the structural study of many other membrane proteins as well.

The BK channel³ has many physiological roles: it controls firing patterns in neurons, modulates the tone of blood vessels, and in some animals is an element of the electrical resonator in the ear. Among ion channels, it has served as a model system because of its remarkable ion-permeation properties⁴ and its accessibility for studies of allosteric control of gating^{3,5}. It is formed as a tetramer of α -subunits expressed from the *Slo* gene⁶, which in the human genome is called *KCNMA1*. Alternative splicing⁷ of *Slo* transcripts result in channels having differing conductance and gating properties, whereas co-expression with various β -subunits results in channels having differing Ca^{2+} sensitivity and degrees of inactivation⁶. Like other members of the 6TM ion-channel family, BK has voltage-sensor domains (VSDs) that confer the primary sensitivity to membrane potential. The BK channel and other members of the *Slo* family α -subunits also contain regulator of conductance for K^+ (RCK) domains in the large intracellular carboxy-terminal region; these confer⁸ the sensitivity to Ca^{2+} and form a 'gating ring'. Unlike most 6TM α -subunits, *Slo* contains an extra transmembrane segment, S0, and has an extracellular amino terminus (Fig. 1a).

Membrane proteins were extracted from HEK293 cells stably expressing Flag-tagged human *SLO*. They were purified by an anti-Flag affinity column, and reconstituted into 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) liposomes with detergent removal by gel filtration. The resulting proteoliposomes were

separated from empty liposomes and free *SLO* protein through a discontinuous gradient centrifugation process (Supplementary Fig. 1a, b). The 125-kDa human *SLO* protein is seen by SDS-PAGE (Supplementary Fig. 1c). The reconstitution was adjusted to yield an average protein content of one to two BK channels per 30-nm proteoliposome.

The function of reconstituted BK channels was assayed using the cationic fluorescent dye JC-1 to monitor K^+ -induced changes in membrane potential^{9,10}. Proteoliposomes loaded with 135 mM KCl were diluted into 5 mM KCl and the red fluorescence of JC-1 aggregates was measured. A subsequent decrease in fluorescence as external KCl concentration was increased indicates K^+ -selective permeability (Fig. 1b).

Iberitoxin¹¹ is a highly selective blocker of BK channels, binding to the extracellular face of the pore. Added to the external solution,

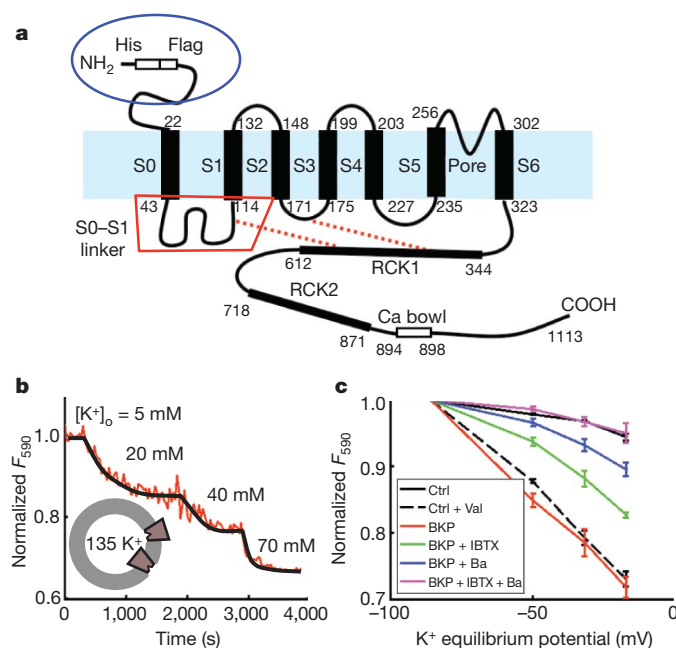


Figure 1 | BK channel structure and specific potassium permeability.

a, Topology and domain structure of the human *SLO* α -subunit of the BK channel. Residue numbers of native human *SLO* are shown; the His and Flag tags add a further 14 residues to the N terminus in our construct.

b, Fluorescence assay of proteoliposome membrane potential. **c**, Normalized fluorescence as a function of calculated potassium equilibrium potential. Signals from empty POPC liposomes (Ctrl) or liposomes in the presence of 1 nm valinomycin, a K^+ ionophore (Ctrl + Val) are compared with those from BK proteoliposomes (BKP) alone or with external addition of the blockers 10 mM Ba^{2+} or 30 μM iberitoxin (IBTX). Error bars indicate s.e.m. ($n = 3-5$).

¹Department of Cellular and Molecular Physiology, Yale University, 333 Cedar Street, New Haven, Connecticut 06520, USA.

iberiotoxin partly reduced the fluorescence signal from the proteoliposomes, as did Ba^{2+} , which blocks the pore from the intracellular side¹². The combination of both blockers reduced the K^+ flux to control levels (Fig. 1c and Supplementary Fig. 2). We conclude that BK channels were inserted in both orientations in the vesicle membranes, with most channels oriented inside out.

Single-particle reconstruction of unstained cryo-EM specimens typically requires the acquisition of 10^4 – 10^5 particle images. Acquiring this many images of protein particles in liposomes is challenging because at most only a few tens of liposomes are present in a typical micrograph, which spans less than $1\ \mu\text{m}^2$ of specimen area. To obtain a uniform, high density of BK proteoliposomes to optimize data collection, we used a two-dimensional (2D) streptavidin crystal as an affinity surface in the cryo-EM specimens. Proteoliposomes, doped with a few copies of biotinylated lipid and osmotically swollen to ensure a spherical shape, were allowed to attach to the crystal (Fig. 2a) before blotting and rapid freezing of the specimen. Low-dose electron-microscope images (Supplementary Fig. 3a) show periodic information from the 2D crystal that can be used as an image-quality reference¹³. For further processing, we computationally removed the crystal information from images (Fig. 2b).

Ideally, a 3D reconstruction would contain an entire proteoliposome, complete with BK channel and spherical membrane. Unfortunately, the variability of liposome size precludes the merging of their images; instead, we fitted and subtracted a model¹⁴ of the membrane contribution to each image and reconstructed the protein particles alone (Fig. 2c). In the RSC method used here, the determination of the angles of orientation of each particle is greatly aided¹⁵ by the spherical vesicle geometry. The apparent position of the particle in a projection image, relative to the vesicle centre, specifies two of the three Euler angles (Supplementary Fig. 4a) within a fourfold

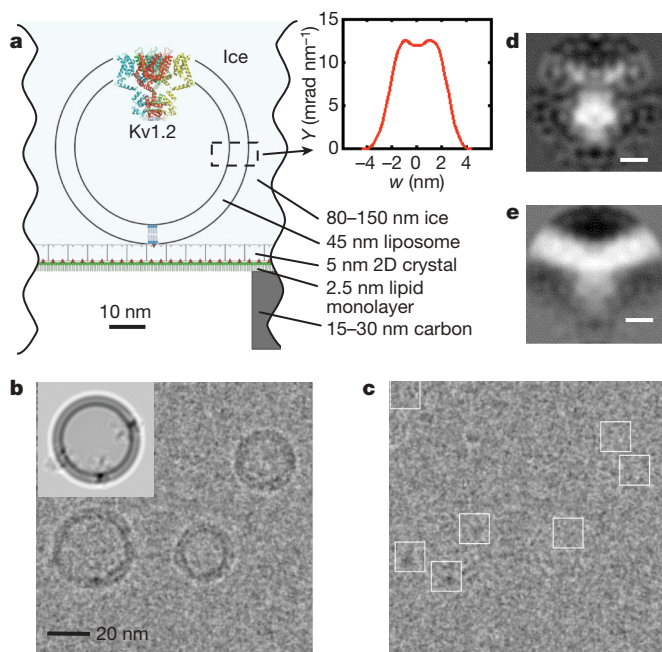


Figure 2 | Cryo-EM specimen and image processing. **a**, Scale drawing of the tethered proteoliposome system. The inset shows the electron-scattering profile of the POPC bilayer. **b**, Electron microscope image with the periodic crystal information removed. The inset shows a simulation in which the membrane profile and three copies of the BK structure are oriented to reproduce the proteoliposome image underneath. **c**, The same micrograph after subtraction of modelled membranes. BK channel particles were selected manually (white boxes). **d**, **e**, Central sections of the 3D reconstruction of BK channels after subtraction of the membrane density, and after a patch of membrane was computationally restored, respectively; scale bar, 5 nm.

ambiguity. Our 3D reconstructions relied on the Fourier-space reconstruction strategy of Grigorieff¹⁶ but with geometric constraints applied to the angular search. The 3D reconstructions imposed C4 symmetry and used subsets of the 8,400 particle images from 644 micrographs. The resulting electron-scattering density map (Fig. 2d) shows a low density in the transmembrane region, as expected from the subtraction of the modelled membranes. A parallel reconstruction, performed using the same angle assignments but with a patch of membrane density restored to each particle image, illustrates the curved membrane (Fig. 2e). The resolution of a reconstruction from the entire data set was estimated to be 1.7–2.0 nm by the Fourier shell correlation (Supplementary Fig. 4b). Differences in vesicle size, and therefore membrane curvature, appear to have little effect on the channel structure (Supplementary Fig. 5).

Two-thirds of the SLO protein sequence forms the large cytoplasmic C-terminal domain; we therefore assign the large particle mass that was usually found external to a proteoliposome to be the C-terminal domains of an inside-out BK channel. With the membrane potential close to zero and free calcium in the nanomolar range, we expect the derived structure reflects a closed channel. Only 3% of voltage sensors are activated⁵ under these conditions.

The transmembrane region of BK, containing the pore and voltage-sensor domains (Fig. 3), is similar in extent to that seen in recent X-ray crystal structures of 6TM potassium channels: the voltage-gated channel Kv1.2 structure shows an open (or possibly

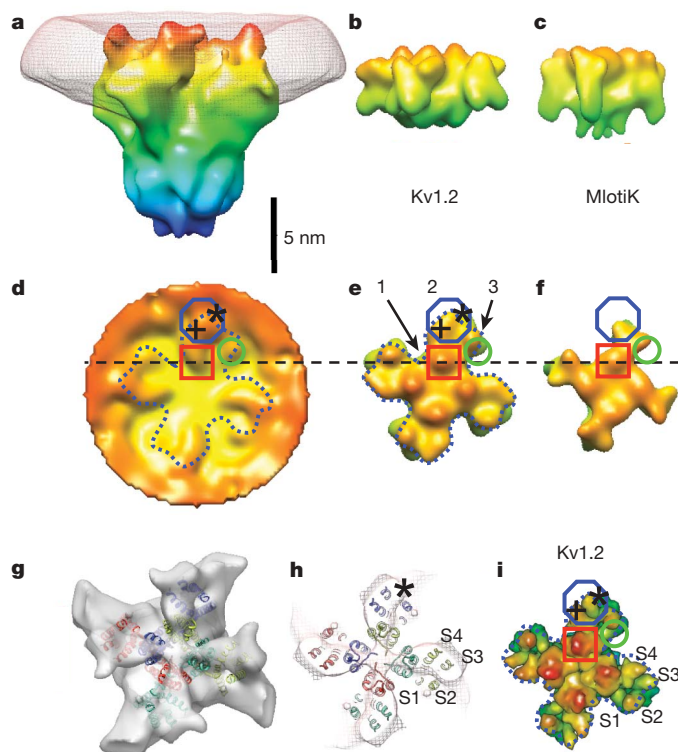


Figure 3 | Structure of the transmembrane region. **a**, Surface rendering of the membrane-subtracted, inside-out BK channel map, obtained from 3,400 images of particles in large vesicles. Superimposed is the membrane density (mesh). Maps were filtered to a resolution of 2.0 nm; isosurfaces are coloured according to the z-coordinate. **b**, **c**, Surface renderings of Kv1.2 and MlotiK X-ray structures, filtered to 1.7-nm resolution for comparison. **d**, Extracellular aspect of the membrane-restored reconstruction of BK. **e**, **f**, Corresponding extracellular views of Kv1.2 and MlotiK. **g**, Extracellular aspect of the membrane-subtracted BK map (solid) with transmembrane helices of the docked Kv1.2 structure superimposed. **h**, Section (2-nm thick) of the membrane-subtracted BK map (mesh) near the membrane centre with the corresponding Kv1.2 helices superimposed. **i**, Surface rendering of the Kv1.2 transmembrane region, but with a resolution of 0.3 nm.

depolarized-inactivated) state^{17,18} whereas MlotiK1, a prokaryotic ligand-gated channel, is seen in its closed state¹⁹. Because the membrane-subtraction process modifies the densities of the protein at the membrane–aqueous interface, it is best to use the membrane-restored cryo-EM map (Fig. 3d) to compare the extracellular face of BK with the other 6TM channel structures. BK shows protrusions corresponding to the turret region of the pore domain (red square) and the S2 helix of the voltage-sensor (green circle). There is good correspondence of these features to both the Kv1.2 and MlotiK1 X-ray structures, and at this resolution there is expected to be little difference in the extracellular face between open and closed states. BK, however, shows a much larger protrusion at the periphery of the VSD (blue octagon). The additional helix S0 and the extracellular ~40 N-terminal residues are expected to give a feature of this size.

In a plane at the centre of the membrane, features in the cryo-EM density map are little affected by membrane subtraction; there the envelope of each VSD is seen to correspond well to the four-helix bundle of the Kv1.2 structure (Fig. 3h). Compared with Kv1.2, the cryo-EM map contains additional density at the VSD periphery (see also Fig. 4e), which can account for the additional S0 helix. This location, which would place it in contact with S2 and S3, is consistent with a recent cross-linking study²⁰. On the other hand, the cryo-EM map does not match the configuration of VSD helices in the KvAP crystal structure²¹, which is thought to reflect a non-native conformation.

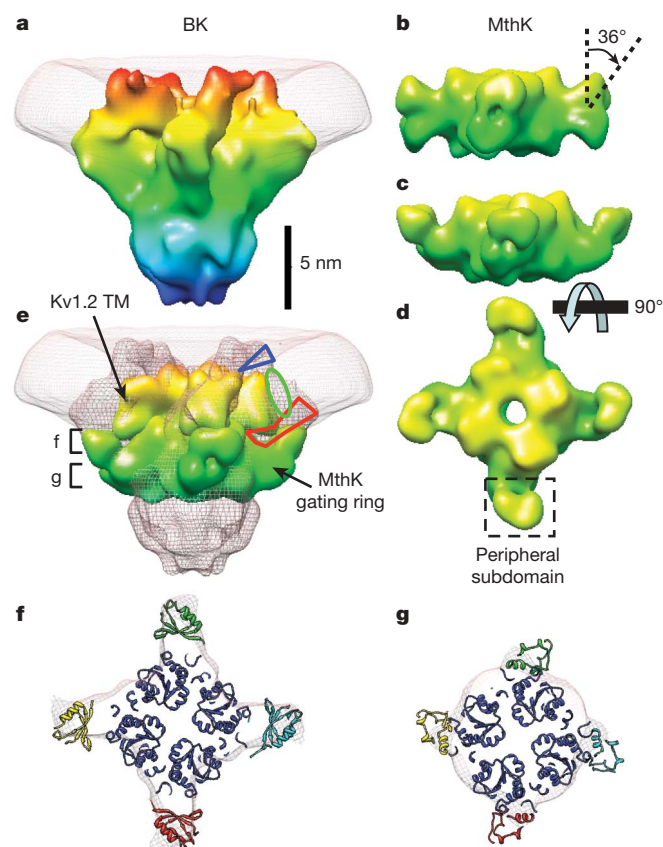


Figure 4 | Structure of the gating ring. **a**, Side view of BK and membrane, rotated 45° about the vertical axis from the view in Fig. 3a. **b**, Surface rendering of the ‘closed’ MthK gating ring filtered to 1.7-nm resolution. **c**, **d**, The MthK gating ring modified by a 36° rotation of the peripheral domains. **e**, View of the BK map (mesh) with Kv1.2 transmembrane region and MthK gating ring docked. All models are coloured according to the z-coordinate. The blue triangle and red polygon indicate possible locations for the N-terminal region and the S0–S1 linker, respectively. The green oval is the proposed location of the S0 helix. **f**, **g**, ‘Top’ views of the BK map and docked MthK gating ring in the sections marked in **e**.

The gating ring, a calcium sensor region formed by RCK domains, is apparent in the cryo-EM map of BK. In the prokaryotic Ca²⁺-activated K⁺ channel MthK, the gating ring consists of eight identical RCK domains. In BK, each α -subunit is thought to contain two RCK domains, yielding a total of eight RCKs in the tetrameric channel complex^{8,22,23}. The inner domain of the MthK gating ring, the structure formed by helices A to F and associated β -strands, is well conserved in nearly all RCK sequences⁸. The RCK peripheral domain, which produces the four protruding regions in the MthK gating ring, is formed by helices G to J and their associated β -strands and is variable among different species⁸. As expected, the well-conserved inner domain of the closed MthK gating ring²⁴ is readily docked into the cryo-EM map of BK (Fig. 4e, f). The periphery of the MthK gating ring does not match the strong continuous density near the transmembrane region of BK. This density can be better matched by tilting the peripheral domains of the MthK gating ring by 36° (Fig. 4c, d). Even so, the tilted MthK peripheral domains contain excess mass compared with the cryo-EM map (Fig. 4g), consistent with the idea that in BK the RCK2 domain is truncated and is about 100 amino-acid residues shorter than RCK1.

The calcium bowl, a high-affinity calcium-binding site that lies after RCK2, might also reside in the gating ring. Alternatively, it could be located below the gating ring in the large mass of density which we assign to the remainder of the protein sequence. This density is of the correct size to encompass the 240 C-terminal residues (Fig. 4e), including the calcium bowl.

The close apposition between the transmembrane region and the gating ring is consistent with functional studies²³ demonstrating the formation of an Mg²⁺-binding site between residues in the transmembrane region and in the gating ring. These residues are Asp 99 in the S0–S1 linker and Asn 172 in S2 of the transmembrane domain, with residues Glu 374 and Glu 399 in RCK1 of the BK gating ring (Fig. 1a). In the MthK channel, the gating ring is connected to the S6 helix in the pore region by a disordered 17-residue linker, leaving a cleft with a maximum width ~1 nm in the open state⁸; the resulting lateral openings allow ions to access the ion conduction pore. A similar linker is expected to be formed in BK channels, but a cleft of this size would not be visible in our map, even though it must be large enough to accommodate the inactivation domain of the β 2 subunit²⁵.

A new single-particle reconstruction technique, RSC cryo-EM, thus provides the first structural model of the Ca²⁺- and voltage-activated K⁺ channel. The disposition of the VSDs of 6TM channels has been a matter of controversy, but we find that the membrane-embedded channel’s VSDs match well with two recent 6TM channel X-ray crystal structures. The calcium-sensing ‘gating ring’ is also visible in the density map, as is the density corresponding to regions unique to this channel protein. The RSC technique has the added advantage that the channel protein is imaged in a membrane environment where channel activity can also be assayed. It should be possible to obtain reconstructions at higher resolution by using more particle images than we used here.

METHODS SUMMARY

Purification and reconstitution of BK protein. Full-length human SLO protein (gi:507922) carrying an N-terminal Flag tag was obtained from an HEK293 cell line²⁶. Protein was solubilized in dodecylmaltoside and purified using a Flag antibody affinity column, where the detergent was exchanged with decylmaltoside before elution with Flag peptide. BK protein was concentrated and added to decylmaltoside-solubilized POPC lipid (decylmaltoside:POPC = 3:1), giving a final protein:lipid molar ratio of 1:5,000. Also present in the lipid mixture was biotinylated dipalmitoyl-phosphatidylethanolamine, at a low concentration (1/3,600 of POPC) calculated to yield about three copies per 30 nm liposome. Gel filtration was used to remove detergent. After the liposomes were concentrated, they were floated on a discontinuous Nycodenz gradient. Protein-free liposomes were found in the 3% Nycodenz band, whereas proteoliposomes appeared at the 5–15% boundary. Lipid concentrations were determined by measuring phosphate using a colorimeter; protein was determined with the Micro BCA Protein

Assay Kit (Thermo Scientific) and the relative fraction of BK in each layer was determined by densitometry of a western blot.

Tethering of BK proteoliposomes and cryo-EM imaging. Two-dimensional streptavidin crystals were grown at room temperature using the procedure described previously¹³. After the 2D streptavidin crystal was transferred to the perforated carbon film, the crystal was incubated with BK proteoliposome suspensions for 10–40 min to allow binding. The sample was blotted at room temperature and immediately fast-frozen in liquid ethane. Samples were imaged at -180°C in a Tecnai F30 microscope at 300 keV with a 30- μm objective aperture and zero-loss energy filtering. The electron dose at the specimen was 1,000–3,000 electrons nm^{-2} . Images were taken at $\times 50,000$ magnification and -2 - to -5 - μm defocus, and recorded on a GIF 2Kx2K UltraScan 1000 FT camera with an effective pixel size of 0.253 nm.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 February; accepted 17 July 2009.

Published online 30 August 2009.

- Raunser, S. & Walz, T. Electron crystallography as a technique to study the structure on membrane proteins in a lipidic environment. *Annu. Rev. Biophys.* **38**, 89–105 (2009).
- Frank, J. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies* (Oxford Univ. Press, 2006).
- Cui, J., Yang, H. & Lee, U. S. Molecular mechanisms of BK channel activation. *Cell. Mol. Life Sci.* **66**, 852–875 (2009).
- Neyton, J. & Miller, C. Discrete Ba^{2+} block as a probe of ion occupancy and pore structure in the high-conductance Ca^{2+} -activated K^{+} channel. *J. Gen. Physiol.* **92**, 569–586 (1988).
- Horrigan, F. T. & Aldrich, R. W. Coupling between voltage sensor activation, Ca^{2+} binding and channel opening in large conductance (BK) potassium channels. *J. Gen. Physiol.* **120**, 267–305 (2002).
- Salkoff, L. et al. High-conductance potassium channels of the SLO family. *Nature Rev. Neurosci.* **7**, 921–931 (2006).
- Fodor, A. A. & Aldrich, R. W. Convergent evolution of alternative splices at domain boundaries of the BK channel. *Annu. Rev. Physiol.* **71**, 19–36 (2009).
- Jiang, Y. X. et al. Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* **417**, 515–522 (2002).
- Chanda B. & Mathew, M. K. Functional reconstitution of bacterially expressed human potassium channels in proteoliposomes: membrane potential measurements with JC-1 to assay ion channel activity. *Biochim. Biophys. Acta* **1416**, 92–100 (1999).
- Reers, M. J-aggregate formation of a carbocyanine as a quantitative fluorescent indicator of membrane potential. *Biochemistry* **30**, 4480–4486 (1991).
- Galvez, A. et al. Purification and characterization of a unique, potent, peptidyl probe for the high conductance calcium-activated potassium channel from venom of the scorpion *Buthus tamulus*. *J. Biol. Chem.* **265**, 11083–11090 (1990).
- Miller, C., Latorre, R. & Reisin, I. Coupling of voltage-dependent gating and Ba^{++} block in the high-conductance, Ca^{++} -activated K^{+} channel. *J. Gen. Physiol.* **90**, 427–449 (1987).
- Wang, L., Ounjai, P. & Sigworth, F. J. Streptavidin crystals as nanostructured supports and image-calibration references for cryo-EM data collection. *J. Struct. Biol.* **164**, 190–198 (2008).
- Wang, L., Bose, P. S. & Sigworth, F. J. Using cryo-EM to measure the dipole potential of a lipid membrane. *Proc. Natl Acad. Sci. USA* **103**, 18528–18533 (2006).
- Jiang, Q. X., Chester, D. W. & Sigworth, F. J. Spherical reconstruction: a method for structure determination of membrane proteins from cryo-EM images. *J. Struct. Biol.* **133**, 119–131 (2001).
- Grigorieff, N. FREALIGN: high-resolution refinement of single particle structures. *J. Struct. Biol.* **157**, 117–125 (2007).
- Long, S. B., Campbell, E. B. & MacKinnon, R. Voltage sensor of Kv1.2: structural basis of electromechanical coupling. *Science* **309**, 903–908 (2005).
- Long, S. B. et al. Atomic structure of a voltage-dependent K^{+} channel in a lipid membrane-like environment. *Nature* **450**, 376–383 (2007).
- Clayton, G. M. et al. Structure of the transmembrane regions of a bacterial cyclic nucleotide-regulated channel. *Proc. Natl Acad. Sci. USA* **105**, 1511–1515 (2008).
- Liu, G. et al. Position and role of the BK channel α subunit S0 helix inferred from disulfide crosslinking. *J. Gen. Physiol.* **131**, 537–548 (2008).
- Jiang, Y. X. et al. X-ray structure of a voltage-dependent K^{+} channel. *Nature* **423**, 33–41 (2003).
- Fodor, A. A. & Aldrich, R. W. Statistical limits to the identification of ion channel domains by sequence similarity. *J. Gen. Physiol.* **127**, 755–766 (2006).
- Yang, H. H. et al. Activation of Slo1 BK channels by Mg^{2+} coordinated between the voltage sensor and RCK1 domains. *Nature Struct. Mol. Biol.* **15**, 1152–1159 (2008).
- Ye, S. et al. Crystal structures of a ligand-free MthK gating ring: insights into the ligand gating mechanism of K^{+} channels. *Cell* **126**, 1161–1173 (2006).
- Zhang, Z. et al. A limited access compartment between the pore domain and cytosolic domain of the BK channel. *J. Neurosci.* **26**, 11833–11843 (2006).
- Bingham, J. P. et al. Synthesis of a biotin derivative of ibertoxin: binding interactions with streptavidin and the BK Ca^{2+} -activated K^{+} channel expressed in a human cell line. *Bioconjug. Chem.* **17**, 689–699 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. Grigorieff and C. Xu for use of the Tecnai F30 microscope and help with data collection. We also thank S. Bian for sharing his BK purification methods, and Y. Yan for cell culture. Image processing and reconstruction made use of the Yale Biomedical High-Performance Computing Center.

Author Contributions L.W. and F.J.S. designed the experiments; L.W. purified and reconstituted the protein, made and imaged the cryo-EM specimens, and performed image processing and reconstruction. L.W. wrote custom software, with some algorithms contributed by F.J.S. L.W. and F.J.S. co-wrote the paper.

Author Information The density maps are deposited in the Electron Microscopy Data Bank under accession numbers EMD-5114 and EMD-5121. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to F.J.S. (fred.sigworth@yale.edu).

METHODS

Membrane potential measurements with JC-1 to assay ion channel activity. 5P,5P,6,6P-tetrachloro-1,1,3,3P-tetraethylbenzimidazolylcarbocyanine iodide (JC-1, Invitrogen) was used to assay ion-channel activity using a method similar to that described previously⁹. BK proteoliposomes loaded with 135 mM KCl were incubated with channel blockers if desired, and diluted into 5 mM KCl solution containing 1.6 μ M of JC-1, giving a total lipid concentration of \sim 6 μ M. The fluorescence signal of the J-aggregates (λ_{ex} = 480 nm, λ_{em} = 590 nm) was monitored as the external K⁺ concentration was increased by the addition of 2 M KCl solution.

The large permeability of BK channels (\sim 10⁸ ions s⁻¹) and the small liposome size means that the membrane potential of a liposome will be established very quickly once a channel opens, as the movement of only about 100 ions is required. We performed the flux assays in nanomolar free Ca²⁺ and at liposome membrane potentials near zero, to reduce the BK open probability⁵ to about 10⁻⁵. Even so, only one functional channel provides a maximal fluorescence signal from a given liposome. The timescale of charging is so much faster than the timescale of redistribution of JC-1 across the membrane (tens of seconds) that the block of BK channels must be very complete if a reduction in the fluorescence response to K⁺ gradients is to be observed. Thus we applied channel blockers at about 1,000 K_d to reduce the net permeability to control levels.

Image processing. The periodic 2D streptavidin crystal information was removed computationally as described¹³, and the liposome membrane contribution was removed using a model based on the average POPC membrane profile from 250 micrographs. The model was obtained using the Hankel transform as described¹⁴. Images of liposomes smaller than 20 nm were not used. BK particle images were manually picked using EMAN boxer²⁷ and the contrast-transfer function parameters were estimated using a homemade Matlab program¹³ from the power spectrum of each micrograph. The picked BK particle images, with crystal and liposome information removed, were used to determine the 3D structure employing constraints based on the spherical geometry of the proteoliposomes. First, two of the three Euler angles (θ and ψ ; Supplementary Fig. 4a) were estimated based on the position in the image of the particle with respect to the liposome centre. To account for uncertainty in the estimate of the particle centre, the two angles were allowed to vary in a range corresponding to a maximum in-plane displacement of the particle centre of 1.8 nm. Then reference images (projection images of the 3D BK map from the earlier iteration) that fell within these ranges were computed, using an angular step size of 3°. The references were sorted according to the similarity (cross correlation coefficient) with the particle image, and tested from the best to the worst match until the predicted position based on the angles of that reference was consistent with the observed position in the cryo-EM image. Then the angles of that reference were assigned as the angles of the BK particle. All possible orientations of the BK channels were included in this data set (Supplementary Fig. 6).

After three cycles of initial search, we refined the structure in the same way except that the absolute value of the correlation coefficient, rather than the correlation coefficient itself, was maximized. This serves to reduce the reference

bias in the final structure²⁸. Finally, the 3D map was constructed with equal weighting of each particle image, using the least-squares Fourier-space algorithm that is used in the FREALIGN program¹⁶. Because the handedness of the reconstruction from projections was not known, we compared the intra-membrane density of Kv1.2 and the BK map to make the assignment.

The first reconstruction was a 3D electron-scattering map of BK channels with membrane removed (Fig. 2d). Then each particle image was modified by re-inserting the image of a patch of membrane in the following way. Based on the assigned Euler angles and the model of the corresponding lipid vesicle, a spherical sector of membrane centred on the particle was defined. The projection of this membrane patch, modified by the contrast transfer function, was then added back to the particle image and a second 3D map was constructed (Fig. 2e). Because the angle assignments and scaling were maintained between the two reconstructions, the difference quantitatively shows the membrane density. Given estimates²⁹ of the internal potential of protein, lipid membrane and water, the contrast between membrane and protein regions is expected to be only about 50% of that between protein and water.

The proteoliposomes used for structure determination ranged from 20 to 60 nm in diameter (Supplementary Fig. 5h). The question therefore arises, whether the membrane curvature in vesicles of different sizes would affect the observed BK structure. To address this, two BK structures were reconstructed from small (20–24.5 nm) and large (24.5–60 nm) BK proteoliposomes respectively (Supplementary Fig. 5). For examination of the ‘extracellular’ channel surface, the threshold of the electron microscope map was first set such that the ‘bare’ membrane thickness was 5.0 nm. In the vicinity of BK channels, the extracellular views of BK from small and large vesicles were seen to be very similar (Supplementary Fig. 5a, b). When the threshold was increased, the features in regions 1–3 (Supplementary Fig. 5d, e) could be identified more clearly both in the small and large liposomes, but between different sizes of proteoliposomes there was still no difference observed in the surface profile of the channel features. A line profile of the isosurface, tracing a path between two opposite VSDs, confirms a lack of height difference out to a distance of 6 nm from the channel axis (Supplementary Fig. 5g). Thus the effect of the membrane curvature on the BK structure at the current resolution appears to be small.

All image processing and other numerical calculations were done in the Matlab programming environment (MathWorks). The docking of density maps derived from the crystal structures of MlotK, Kv1.2 and the MthK gating ring (Protein Data Bank codes 3BEH, 2A79 and 2FY8) was performed manually using Chimera³⁰.

27. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
28. Stewart, A. & Grigorieff, N. Noise bias in the refinement of structures derived from single particles. *Ultramicroscopy* **102**, 67–84 (2004).
29. Grigorieff, N., Beckmann, E. & Zemlin, F. Lipid location in deoxycholate-treated purple membrane at 2.6 angstrom. *J. Mol. Biol.* **254**, 404–415 (1995).
30. Pettersen, E. F. *et al.* UCSF Chimera – a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

ERRATUM

doi:10.1038/nature08359

MicroRNA-mediated switching of chromatin-remodelling complexes in neural development

Andrew S. Yoo, Brett T. Staahl, Lei Chen & Gerald R. Crabtree

Nature 460, 642–646 (2009)

In the print issue of this Letter, Fig. 3 was incorrectly printed as a black and white image. The correct image is shown below.

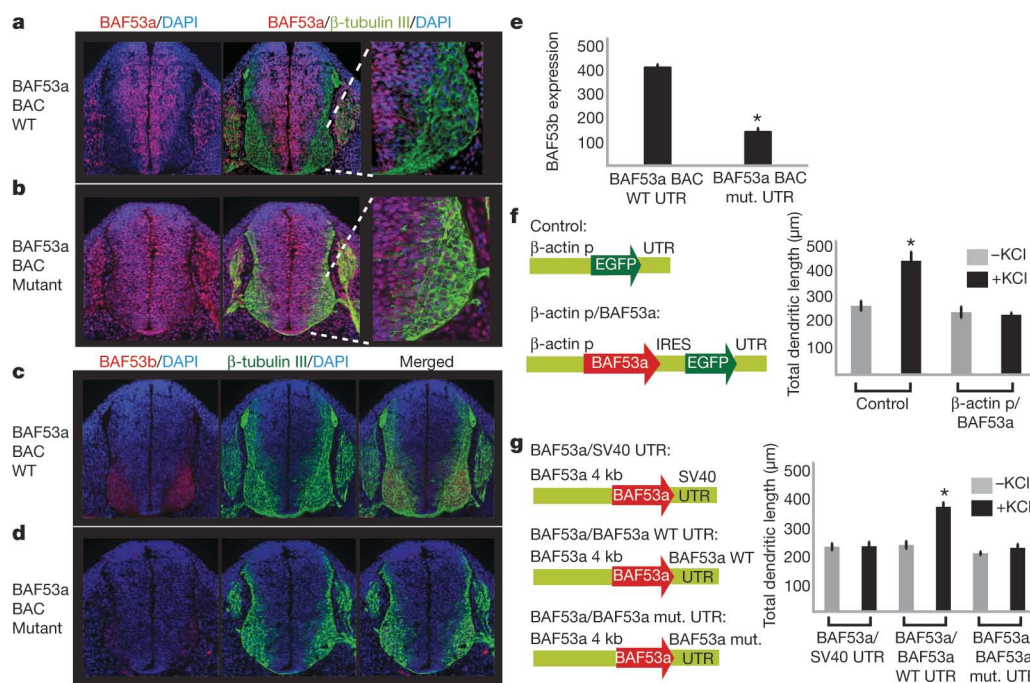


Figure 3 | BAF53a repression is essential for activity-dependent dendritic outgrowth in neurons. **a**, Normal downregulation of BAF53a in post-mitotic neurons in transgenic embryos with wild type BAF53a BAC. The rightmost panel shows the lower-right quadrant of the neural tube. **b**, Persistent expression of BAF53a in neurons seen with BAF53a BAC containing point mutations in the miRNA-binding sites. **c**, Normal expression of BAF53b (red) in β-tubulin-III-positive (green) neurons in transgenic embryos with wild-type BAF53a BAC. **d**, Reduced BAF53b expression with persistent expression of BAF53a in neurons. **e**, Quantification of BAF53b expression: ratio of BAF53b level (arbitrary units) and β-tubulin-III-positive neurons. Average values are from eight sections of the neural tube. Error bars, s.e. * $P < 0.005$, Student's *t*-test.

f, Constructs to overexpress BAF53a in cultured hippocampal neurons and quantification of dendritic outgrowth of GFP-positive neurons upon stimulation using KCl. The average values are from five individual coverslips from two independent experiments, with each coverslip containing 50–100 scored neurons. Error bars, s.e. * $P < 0.005$, Student's *t*-test. p, promoter; IRES, internal ribosome entry site. **g**, Schematic diagrams of BAF53a expression constructs using different 3' UTRs and quantification of dendritic outgrowth of transfected neurons upon stimulation using KCl. In independent experiments, we found that the 4-kb upstream region of BAF53a (illustrated) was sufficient to drive expression of GFP reporters that could be repressed by endogenous miR-9* and miR-124. Error bars, s.e. * $P < 0.001$, Student's *t*-test.

MASTER your future

The University of Pennsylvania's innovative, interdisciplinary **Master of Environmental Studies** program will give you the knowledge to solve complex environmental problems at the local and global levels.

Work alongside some of the brightest minds in the field as you focus on Environmental Policy, Resource Management, Environmental Biology or Environmental Health, or center your studies around a specific environmental field that matches your interest.

Get involved: Check out the Philadelphia Global Water Initiative and Greenworks Philadelphia.

Visit our web site or call to find out how Penn's Master of Environmental Studies program can help you make a difference.

GreatThinking.org | 215.898.7326



College of Liberal & Professional Studies
Penn LPS

NW184753R

SPOTLIGHT ON PHILADELPHIA



Don't miss the intoxicatingly good job opportunities in *Nature* each week and on naturejobs.com

naturejobs

AACR American Association
for **Cancer Research**

AACR Membership

Members Dedicated to Conquering Cancer
Through Research, Education and Training,
Communication and Collaboration



www.aacr.org/membership

NW184570A

FOX CHASE CANCER CENTER

Where Basic Cancer Research Leads to Advanced Cancer Medicine

Nestled on a leafy hilltop in Philadelphia, Fox Chase Cancer Center is an independent National Cancer Institute-designated Comprehensive Cancer Center with over a century's worth of experience in the study and treatment of the disease.

Indeed, Fox Chase was known for translational research long before the phrase became a popular buzzword. How else to depict the work of Nobelist Baruch Blumberg, discoverer of the hepatitis B virus and its vaccine? What better term describes the discovery of ubiquitin-mediated protein degradation by Fox Chase researchers, including Irwin A. Rose, which also earned them a Nobel Prize and opened up a new avenue for drug research?

From the discovery of vitamin B12—to the development of the first transgenic and chimeric animals—to the discovery of the SCID mouse, Fox Chase embodies an impressive scientific legacy.

Scientific Research Programs

Fox Chase supports approximately 100 researchers and physician-scientists through an NCI Cancer Center Support Grant, which also funds an array of advanced research facilities to meet their needs. Together, our researchers have created a collaborative atmosphere for research, where new ideas are cherished and explored.

Our six core research programs include:

- Epigenetics and Progenitor Cells
- Immune Cell Development and Host Defense
- Cancer Genetics and Signaling
- Cancer Prevention and Control
- Molecular Translational Medicine
- Women's Cancer

Keystone Programs

With so many researchers working in close proximity to Fox Chase's distinguished medical staff, it is no surprise that Fox Chase is deeply invested in translational research.

In February 2008, Fox Chase launched a suite of innovative team-based cancer research initiatives. At the heart of every Keystone Program is a self-organized group of scientists and clinicians utilizing the strengths of our core research programs in order to hasten medical progress against cancer. Each Keystone Program was vetted through a comprehensive external peer-review board before receiving a five-year, multi-million-dollar start-up grant with money raised through private philanthropy.

The Keystone Programs include:

- Epigenetics and Progenitor Cells
- Personalized Kidney Cancer Therapy
- Personalized Risk Assessment and Prevention
- Blood Cell Development and Cancer
- Head and Neck Cancer

SPORE in Ovarian Cancer

Fox Chase was selected to lead an NCI grant for a Specialized Program of Research Excellence (SPORE) in prevention, diagnosis and treatment of ovarian cancer. The Fox Chase-Penn SPORE has developed advanced ongoing projects in epigenetics, biomarker discovery and validation, and targeted therapeutics.

Institute for Personalized Medicine

Decades of experience and research in cancer care tells us that no two cancers are alike: some respond well to a particular therapy, while others, seemingly identical, do not respond at all. In order to make the "one-size-fits-all" approach to cancer therapy a thing of the past, Fox Chase has brought together its extensive biosample repository and renowned Phase I Clinical Trials Program to create the Institute for Personalized Medicine (IPM).

The IPM will make it possible to pair individuals with clinical trials for emerging targeted therapies, thereby accelerating the rate at which new, powerful drugs become part of the clinical arsenal.

Women's Cancer Center

This fall, Fox Chase opens the new Robert C. Young, M.D., Pavilion, and with it the Fox Chase Women's Cancer Center. The new center ties together Fox Chase clinicians who specialize in breast and gynecologic cancers, and will work closely with the researchers all across campus, including the Women's Cancer Program, the Keystone Programs, and the IPM.

Fox Chase Cancer Center Buckingham

Open since mid-July, this free-standing facility in central Bucks County (about 19 miles from main campus) houses the latest in radiation technology, a field in which Fox Chase researchers are distinguished as both innovators and clinical experts. Among the radiotherapy tools available are the CyberKnife Robotic Radiosurgery System and a Trilogy Linear Accelerator with Rapid Arc.

Lab Scientist, Breast Cancer

Fox Chase Cancer Center seeks a nationally/internationally recognized laboratory scientist in the field of breast cancer research. Generous start-up package and new state-of-the-art laboratory space will be offered.

The ideal candidate is an active senior scientific researcher at the associate or full professor level with a proven history of peer-reviewed grant funding and a strong record of breast cancer translational research. Must have Ph.D. and/or M.D.

The laboratory scientist will work closely with clinicians and researchers who specialize in breast and gynecologic cancers in Fox Chase's Women's Cancer Center, part of the newly opened Robert C. Young, M.D., Pavilion. With some of the nation's top clinicians and researchers working together, Fox Chase is able to deliver powerful and innovative therapies. The Women's Cancer Center offers a comprehensive approach to cancer care, from screening and family risk assessment to pioneering treatment options, including access to clinical trials.

Fox Chase Cancer Center is one of the leading freestanding cancer research and treatment centers in the U.S.

Founded in 1904 in Philadelphia as one of the nation's first cancer hospitals, Fox Chase became one of the first institutions to be designated a National Cancer Institute Comprehensive Cancer Center in 1974.

Fox Chase. Perfecting the Science of Compassionate Cancer Care.

Please send résumés to:
Jeff Boyd, Ph.D.
Chief Scientific Officer
Fox Chase Cancer Center
333 Cottman Avenue
Philadelphia, PA 19111
Jeff.Boyd@fccc.edu

EEOC

NW183801R

www.naturejobs.com

NEWS

Business skills for postdocs

The Keck Graduate Institute in Claremont, California, has established the first US professional master's programme offering business and industry training specifically to postdoctoral fellows. The programme builds on the Professional Science Master's degree, an increasingly popular business-skill building option offered to graduate students.

The institute's president, Sheldon Schuster (pictured), says that the programme aims to respond to industry complaints that academic postdocs often don't understand the corporate culture. A pilot programme will begin this September with four or five students taking existing business courses at Keck. By January 2010, Schuster expects up to two dozen openings as the institution customizes the programme to meet postdoc-specific needs, such as developing industry projects that tackle more complicated scientific questions than graduate students address.

The institute was the first to offer a degree aimed at producing PhD students with business skills. Now there are more than 120 such programmes across the United States.

Many in industry are eager to couple the postdoc's scientific sophistication with business savvy. "For technical companies like us, there is a lot of value in educating people who have demonstrated strong scientific depth with some business skills," says Jim Widergren, corporate vice-president of Asia Pacific and Latin American operations at

Beckman Coulter, a diagnostic biotechnology company based in Fullerton, California.

Joseph Panetta, president and chief executive of BIOCOM, a biotech industry organization

based in San Diego, California, says that BIOCOM created a two- to three-month programme — dubbed the Life Sciences Immersion Program — last year to offer industry skill-building and networking opportunities to postdocs. "Our idea is to give postdocs a bridge into the world of industry," he says. The economic

downturn put the programme on hold, but Panetta hopes to launch it within the next year.

Some postdocs are likely to scoff at yet more training, but those having difficulty securing an academic post may appreciate a new opportunity. "There is going to be a very targeted population of postdoctoral scholars for which Keck's focused industry programme will be a welcome option," says Cathee Johnson Phillips, executive director of the National Postdoctoral Association in Washington DC. She says that there is a growing trend among postdoctoral students to acquire the skills necessary to compete for jobs outside of academia.

Schuster emphasizes the importance of exposing postdocs to all the career options available, especially given the paucity of academic posts compared to the large number of postdocs. "I would be shocked if more professional master programmes specific to postdocs don't pop up in the future," he says. ■

Virginia Gewin



T. ZASADZINSKI

IN BRIEF

Foreign admissions fall

Offers to prospective international students by US graduate schools have fallen by 3% since last year, the first decline since 2004, according to a report by the Council of Graduate Schools in Washington DC.

More than half of the roughly 250 institutions that responded to the council's survey reported offering fewer places to international students. The countries most affected were India and South Korea, which fell by 16% each. China bucked the trend with a rise of 13%.

Nathan Bell, the council's director of research and policy analysis, attributes the declines to a reduction in recruiting efforts abroad because of the recession.

Data manager for Europe

The European Molecular Biology Laboratory's European Bioinformatics Institute, based in Cambridge, UK, is expanding its data storage and handling capacity to become the hub of the European Life Sciences Infrastructure for Biological Information (ELIXIR) initiative.

ELIXIR is funded by the European Commission and individual federal governments, non-profit organizations and agencies across Europe and aims to create a way to manage and store data from the life sciences and information from thousands of labs. Its establishment could create new positions in bioinformatics and information technology.

IT sector takes a hit

The global recession has hit the information and communications technology industry hard. But research and development in the sector is performing solidly, according to a report issued last month by the Organisation for Economic Co-operation and Development. The organization, which collects economic and science-policy information, among other responsibilities, represents 30 nations worldwide.

The report found that venture-capital investment in the sector has markedly slowed since mid-2008. But it continues to flow into clean technologies in the sector, an area that should continue to receive a major share of venture capital, the document predicts.

POSTDOC JOURNAL

Finding the perfect match



Applying for a faculty job is surprisingly analogous to dating. I screen job adverts and respond to those that look like good matches. If a search committee decides that I'm an attractive candidate, it will suggest that we meet face-to-face to get to know each other.

As I prepare my interview outfit and my research talk, I wonder when I'll no longer have to apply to search committees that never call or e-mail me back. I daydream that I'll soon be out of the job market and in a committed relationship with a department that appreciates my research

and doesn't have unrealistic teaching expectations.

During my interview I build a rapport with potential colleagues as we share our research interests and future aspirations for the department. We explore whether we might become better scientists by working together. Our discussion might become more intimate: how do I feel about collaboration? How many graduate students would I like to work with?

Light-headed, I return home and wait for news. I restrain myself from serially checking e-mail or from hovering by

the phone. Time passes and I wonder why the search committee hasn't called. Perhaps it is busy, or on holiday, or maybe its mother is ill. Would it respond well to a music mix tape? Maybe it's just not that into me. As a distraction, I throw myself into other job applications. Someday soon I'm sure I'll meet that perfect search committee — and we'll know for certain that we are meant to be together. ■

Julia Boughner is a postdoc in evolutionary developmental biology at the University of Calgary, Canada.



EYEING THE UNDERDOG

Can Philadelphia's biotechnology industry absorb the jobs lost from pharmaceutical companies? **Kerry Grens** investigates.

Nestled along the northeastern corridor between the giant hubs of US government and finance, Washington DC and New York, Philadelphia can often go overlooked. Its cultural claims to fame are modest: the cheesesteak, *Rocky* and the Liberty Bell. But its life-sciences industry is not so humble. Located within the city and its suburbs, which extend from Princeton, New Jersey, to Newark, Delaware, is one of the largest bio-business clusters in the country. This year, economists at the Milken Institute — an independent economic think tank in Santa Monica, California — ranked the region as runner-up for the strongest biosciences industry, second only to Boston, based on the number of people it employs, its innovation and the revenue its businesses earn. How long it will keep that distinction, however, is uncertain.

Tens of thousands of people in the region still earn their income from pharmaceutical giants such as Merck, Wyeth, AstraZeneca and GlaxoSmithKline, not to mention from dozens of biotech companies, contract-research organizations, generic-drug makers and academic institutions. For years, those big drug makers provided a refuge for researchers and others from the merciless grind of pursuing tenure in academia and from the economic uncertainty and shoestring budgets of biotech. The region's pharmaceutical firms, with their sleek

suburban campuses, handsome salary-and-benefit packages and unmatched research resources, have traditionally opened their arms wide to scientists from around the world. But those halcyon days may be waning, thanks to economic forces that many contend are irretrievably changing the landscape.

"It was a lot of fun," says McHardy Smith, who for 20 years worked in the ion-channels group at Merck in Rahway, New Jersey. In October last year Smith was laid off as part of the company's restructuring. "I felt great sadness at losing what I had enjoyed as a great job," Smith's severance package was generous, but it has not been able to provide him with a seamless transition to his next job. The day after he was laid off, Smith received a call from a recruiter. "I thought, this will be easy," Smith says. But a position never materialized. "It's been nine months — and I have yet to have anyone say, 'here's your next job.'"

Analysts expect thousands of Smiths — accomplished PhDs at pharmaceutical companies — to be jobless in the near future. As is the case in many other industries worldwide, outsourcing and the recession are taking their toll on domestic job opportunities. And mergers and acquisitions also could deal a severe blow to science jobs

in the region, says Thomas Morr, president of Select Greater Philadelphia, an economic development group. "We're all nervous about that, I think," he adds.

Consolidation

Within weeks of each other this summer, shareholders of Wyeth overwhelmingly approved the company's acquisition by Pfizer, and shareholders of Merck and Schering-Plough voted with similar gusto

to approve their companies' merger. With the exception of Pfizer, the companies all have a substantial presence in the Philadelphia region. The combinations are expected to infuse zest into Pfizer's and Merck's drug pipelines at a time when they face the

dismal future of losing patent protection on such money-makers as the statin Lipitor (atorvastatin) for high cholesterol and Singulair (montelukast) for asthma.

The changes will also have less positive consequences for the region. "We expect a 15% reduction [in our workforce size] from the merger," says Trish Maxson, vice-president of human resources at Merck Research Laboratories. The company has not disclosed which jobs would be lost. Maxson says that the company currently needs scientists, but her quest to fill those positions with external applicants is limited for now.

"The biggest challenge is to not let these people move out of the area."

— Barbara Schilberg

B. KRIST/CORBIS

"We're looking to the Schering talent to see how we can utilize it," she says.

Pfizer and Wyeth have been less forthcoming about job cuts. Wyeth spokesman Douglas Petkus says that it is "premature to discuss any specific changes related to personnel or facilities". Industry advocates say that job losses are on the horizon. "The pharmaceutical industry is consolidating and has been laying off employees in very large numbers," says Ellen Derrico, chair of the life-sciences network of the Greater Philadelphia Senior Executive Group. Derrico cites thousands of lay-offs in the past year at GlaxoSmithKline, Wyeth, Johnson & Johnson, Merck, AstraZeneca and Teva. "I think it will be felt permanently," she says.

"The biggest challenge is to not let these people move out of the area," says Barbara Schilberg, chief executive of BioAdvance. Her organization is one of several life-science investment groups formed by the Commonwealth of Pennsylvania several years ago using money that states had received as part of a massive settlement with tobacco companies. To date, BioAdvance has injected \$16 million into 26 start-up companies and 17 pilot investments. Schilberg proudly says that most of those companies have survived and raised more than \$300 million in venture capital, and six firms were acquired for a total of \$650 million. Schilberg would like to see that new biotech base absorb whatever job losses result from the big pharmaceutical mergers. "In my ideal world, we would rearrange the pieces of the jigsaw puzzle," she says.

Rearranging the pieces

Daniel Skovronsky, chief executive of Avid Radiopharmaceuticals, a recipient of BioAdvance's seed money and one of the first tenants of the University City Science Center's new building, leads a visitor on a tour of his new space: brightly painted walls of orange and gold, a bank of cubicles, a chemistry lab and a vacant clean room. "It's just an empty room," Skovronsky says, "but in my imagination it will be bustling with



Positive: Daniel Skovronsky and Barbara Schilberg.

activity." That's because Avid is expanding. The company's current facility is the third upgrade in Avid's four-year life.

The firm produces imaging agents that tag amyloid plaques in the brain and could be used to diagnose Alzheimer's disease. Skovronsky says that the company's late-stage clinical trials are going well, and he would like to apply the technology to other diseases such as Parkinson's. For him, it's been a good hiring season. "The industry is not as healthy as it once was, so that means there's a better pool of talent," he says. It also means that people may be more willing to grab a lower-paying, less-stable job at a small biotech company than to hold out for one at a large pharmaceutical firm.

Schilberg says that flexibility is the key to landing a job in the current competitive job market. "People need to think outside the box. A full-time job may be the wrong box to squeeze into," she says. Scientists who are willing to consult for multiple companies and gain management or business skills will have a better chance of finding work. It is also often unlikely that they will find a job in a large pharmaceutical firm to simply continue the research they did as a doctoral student.

Smith agrees. He says that he is casting a wide net for his job search that includes contract-research organizations and non-profit advocacy groups. "I'm going to focus on the smaller biotechs because I don't believe that big pharma is going to do a majority of their basic research in-house," Smith says. Indeed, Merck has said it plans to derive 25% of its basic research



through external collaborations. "We think the downsizing of research and development at some of these firms will lead to the creation of some start-up companies," says Morr. The only problem is access to funding — be it venture capital, independent investors, state funds, federal funds or any variation thereof.

The gatekeepers

Skovronsky opens his laptop and pulls up a bar graph of Avid's employment history. "You can see here we started in July 2005, just four years ago, with, well, one employee," he says, pointing to the screen. Every few months the upward trend levels off, then shoots upwards again for a few months. The growth phases occur right after Skovronsky has secured another round of financing for the company. "When the outlook for funding looks uncertain," he says, "the first thing to do is stop hiring."

Venture-capital groups and other investors hold the key to biotech growth. And one of the drawbacks to being in Philadelphia — unlike being in investor hubs such as New York, Boston, and San Francisco — is less proximity to that capital. With venture-capital investments down nationwide, companies are being forced to wait longer to renew or initiate funding. Although the

"The industry is not as healthy as it once was, so that means there's a better pool of talent."
— Daniel Skovronsky

University City Science Center is home to more than 100 companies and comprises 15 buildings at 97% occupancy, the newest structure that opened last year is just 40% full. "In a more favourable economic environment, we'd be leased out already," says Stephen

Tang, the centre's president.

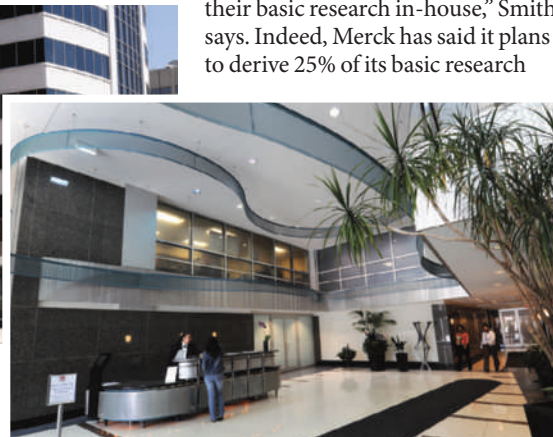
But unlike the changes in the pharmaceutical industry, the scarcity of venture capital is considered a temporary phenomenon, and analysts expect it to recover along with the general economy. Mickey Flynn, the president of Pennsylvania BIO — the state's trade association for the biosciences — is not discouraged. He is pursuing the development of an incubator science centre in Chester County, southwest of Philadelphia. "We hope to generate a lot more innovation and bring some innovation that is in the larger companies out much more quickly," Flynn says.

Schilberg agrees that it is not an easy time for small companies, but she is confident that the number of innovative minds in the region will help to secure their survival. "Chaos always created solutions and creativity," Schilberg says, "and that's what we're doing."

Kerry Grens is the senior health and science reporter at WHYY radio and television station in Philadelphia.



Philadelphia's University City Science Center is home to more than 100 companies.



The pet

An exercise in control.

Robert W. Janes

"The top of the range House Keeper 104 model is an excellent choice!" enthused the salesman. "The best in CAT technology."

"A cat?" queried Leo, with minimum understanding but maximum sneer. His mates had suggested this, and he certainly had the money after all. He could see advantages in having a House Keeper, an ethereal Internet-connected machine that would organize all the boring things in his life, but he had no wish to reveal he wanted this product so that he could show off.

"That is C-A-T," the salesman said, losing the edge off his enthusiasm. "Computer Assisted Thinking. The NNs will anticipate your every need. You'll be..." He halted, as that question mark had reappeared.

"NNs?"

"Neural networks," said the salesman, at a measured pace ensuring his buyer might understand the words at least. He contemplated whether to tell him of the training process by which the Keeper's heuristic parallel processors would assess the best ways to match Leo's needs and wishes. Of the testing phase that would subtly analyse Leo's responses to the Keeper's work, assessing whether they were within anticipated boundary conditions. How, with testing complete, the Keeper would ensure Leo's simplest needs were catered for, giving him freedom to lead an active, carefree life.

He decided on: "This Keeper will take all your troubles away," and left it at that. Life was too short.

"So how do I 'talk' to this Keeper?"

"As naturally as you wish: it will respond almost as if you were talking to a human, but without any emotions."

"I'll call it Cat." The salesman bit his lip. "Seems simple," concluded Leo with a dismissive gesture.

And so do you, thought the salesman.

Leo had been surprised. The voice was female, sultry yet dispassionate, divorced from emotion, a real dichotomy. Two months had gone by. He'd only passingly followed the instructions on how to guide Cat into the correct responses to his wishes. Training had proved difficult with the high demand of interpretation necessary to complete that phase, but Cat was handling all the domestic chores effortlessly, behind the scenes. Somewhere, NNs coped.

"Cat," commanded Leo one evening.



JACEY

"Yes Leo, what do you want?" was the detached response.

"Dinner." He always had his meals brought in from a small home-supply catering company.

"You usually have that at 7.30 and it is now 6.30," came the impartial reply. This was outside normal parameters.

"Now!" yelled Leo. Somewhere, NNs registered. Leo had decided to go out earlier than usual that night and he was going to have a meal, even though he wasn't really hungry.

Their eyes had met across the full distance of the bar. There was something magnetic about Leo. Women were his, but none lasted long in his company. They were thrown away, discarded when his interest quickly waned, but well before theirs had diminished for him. Cat had already included 'company' when in training, and the fit to this action had been excellent in the testing phase; perfectly inside expected limits. So when Bea walked in with Leo, the music was already playing, the wines from his cellar, red and white, were his established choices, the bath was filled, and a relaxing scent wafted through the air filters. Cat had seen to all of that, the timing of return was well within bounds, the water temperature was perfect, the probes indicated as such.

Then he started.

"Cat." His tone, always one of disdain, was even deeper than usual. He really wanted to impress with his power; he was showing off.

"Yes Leo, how can I assist?"

"The bath is too cold, fix it." Before Cat could reply, he went on, "and the wine's not

what I wanted" — he had always had those wines — "and that music — change it, this instant!"

Somewhere NNs bristled. Leo dictated. Cat complied. The bath, half a degree warmer. The wine, change of vintage. The music, same band, different album. I'm in command thought Leo, while Bea looked on and wondered — could she get inside this man's mind?

Bea was different. The one night stretched to two, three, more. Somewhere, NNs waited. Then one day the door closed and it was over. Bea had met only arrogance in her attempts to find Leo, and now it was too late. She had left him.

Leo had tried but Bea's vidi-link was always blocked, so Cat had replied. Now Bea had left, headhunted for a job on the other side of the planet, so Cat had told him.

Leo hardly went out much any more; Bea was not around, so why bother?

"I have got your favourite meal on order tonight Leo. It will be here in five minutes." The music was gently playing in the background; the wine was chilled; scent wafted through the air, as Cat added, slowly, "I know you will like it." Was that intonation in Cat's voice? Leo could not be sure.

"Thank you Cat," replied Leo. The lion was tamed by the Keeper. He was her pet; well within expected limits.

Somewhere, NNs purred.

Bob Janes is a senior lecturer in the School of Biological and Chemical Sciences, Queen Mary University of London, where he researches in the areas of structural biochemistry and bioinformatics.

Join the discussion of Futures in Nature at <http://tinyurl.com/kkh3kt>